The role of chromatin architecture in regulating *Shh* gene during mouse limb development

D I S S E R T A T I O N

zur Erlangung des akademischen Grades
d o c t o r r e r u m n a t u r a l i u m
(Dr. rer. nat.)

im Fach Biologie

eingereicht an der Lebenswissenschaftlichen Fakultät

der Humboldt-Universität zu Berlin

von

Christina Paliou

Präsidentin der Humboldt-Universität zu Berlin
Prof. Dr. -Ing. Dr. Sabine Kunst

Dekan der Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin
Prof. Dr. Bernhard Grimm

Gutachter/innen

1. Prof. Dr. Ana Pombo
2. Prof. Dr. Stefan Mundlos
3. Prof. Dr. Leonie Ringrose

Tag der mündlichen Prüfung: 30th September 2019

1 INTRODUCTION

1.1 Gene regulation in development

As the era of whole genome sequencing began, the results have revealed that morphological complexity does not correlate with the number of genes of each organism. For example, the nematode worm *Caenorhabditis elegans* possesses approximately 20'000 protein-coding genes (C. elegans Sequencing Consortium 1998), while a similar number is found in the human genome (Lander 2001; Ezkurdia et al. 2014). One of the biological mechanisms to develop a complex body plan with different tissues and organs out of the same repertoire of genes is differential regulation of gene transcription. Genes are regulated by *cis*-regulatory elements (CREs) that dictate the level of transcription, in particular when and where a gene is active during development. CREs largely outnumber protein-coding genes and are found at varying distances from the genes they control. Moreover, they often control the expression of more than one target genes, while one gene can be contacted by different CREs in a spatiotemporal specific manner. However, the mechanisms by which CREs identify their target genes and interact in the three-dimensional (3D) space of the nucleus are less understood. Recent technological advances, e.g. proximity ligation-based assays, have offered a better view of how genes and their regulatory partners interact in 3D in each cell and how perturbation of this tight regulation can lead to gene misexpression and disease.

1.1.1 *Cis*-Regulatory Elements

CREs are non-coding stretches of DNA that regulate gene transcription either by promoting or inhibiting it. CREs include promoters as well as silencers, insulators and enhancers that are located either in the vicinity of genes or at greater distances. Among these, promoters and enhancers are the best defined and most studied DNA sequences in relation to gene regulation.

1.1.1.1 Promoters

It is currently thought that gene transcription constitutes one of the most intensely regulated steps of gene expression and occurs in three steps; initiation, elongation and termination. Transcription initiation occurs at promoters, which consist of a gene core promoter and an upstream proximal promoter region. At the 5' end of genes the transcription start site (TSS) is embedded within the core promoter. The core promoter is a short sequence of 50-100 basepair (bp) that functions as a recruiting and binding platform for the transcription machinery or pre-initiation complex (PIC). PIC comprises RNA polymerase II (Pol II) and general transcription factors (GTFs) which are in total

more than 85 polypeptides and sometimes assembled in different combinations (Levine et al. 2014). Although core promoters exhibit only basal transcriptional activity, they are suggested to bear specificity through their sequence but also through the combination and activity of bound transcription factors (TFs) (Zabidi et al. 2014; Zabidi and Stark 2016). As such, core promoters are divided in different types based on their initiation pattern, sequence and motif variability, chromatin conformation and gene function: core promoters of terminally differentiated cell-specific genes, of housekeeping genes or developmental genes (Haberle and Stark 2018).

Finally, promoter proximal regions are located immediately upstream of the core promoter and their function is to drive transcription activation autonomously. As they contain transcription factor binding sites (TFBS), they recruit sequence-specific TFs which participate in the PIC assembly. Interestingly, the fact that proximal promoters exhibit common properties to enhancer elements (Section 1.1.1.2) has raised the question whether they differ from one another and has been a topic of extensive discussion (Haberle and Stark 2018; Medina-Rivera et al. 2018).

1.1.1.2 Enhancers

DNA sequences known as enhancers are crucial players in gene regulation. The first "enhancing" element was described when a 72 bp SV40 viral sequence, cloned next to rabbit *β-globin* and transfected in cultured monkey kidney cells, was able to boost the gene's transcription over 100-fold independently of its orientation and distance to the gene (Banerji et al. 1981). Soon, enhancers were also found to have cell-type-specific preferences (de Villiers et al. 1982) and to bind transcription factors (TFs) in order to control transcription (Lee et al. 1987). After decades of research, the definition of enhancers did not change significantly. Most metazoan enhancers are up to ~ 1 kb long sequences containing binding sites for TFs and cofactors, they increase transcription at target gene promoters and act independently of orientation and distance to the TSS of the target gene. The complex gene regulation in metazoans is achieved by assigning multiple enhancers to single genes, but also by accumulating various TF binding sites on the enhancers. Therefore, the same pool of enhancer elements confers variability in the transcriptional output of one gene by integrating signal from different regulatory and tissue-specific pathways.

1.1.2 Methods for identification of enhancers

Despite the important role of enhancers, it is not always straightforward to identify them due to multiple reasons reviewed in (Pennacchio et al. 2013). First, their relative position to the target gene can be upstream, downstream, intergenic, intronic and even, exonic. Enhancers do not always act on the closest gene, but they can bypass it and regulate a more distant one, located even 1 Mb away

(Lettice et al. 2003). Therefore, it is challenging to associate one gene with its long-range acting enhancer(s). Second, enhancer sequences do not bear any common characteristic binding motifs, in contrast to the known TATA box of some core promoters. Finally, testing enhancer activity in a reporter assay can also be arduous, as it depends on tissue-specific TF availability, thus requiring prior knowledge in order to assess their function. For these reasons, various methods to identify enhancers have been developed over the past decades, yet most of them are reliable when used in combination.

The traditional way of identifying enhancers is based on sequence conservation, which is achieved by comparing the genomes of different organisms (Boffelli et al. 2004). This strategy is based on the assumption that functional regulatory elements are evolutionary conserved because changes in their sequence are likely deleterious to their normal function (Consortium 2002; Chiaromonte et al. 2003). For instance, comparison of the orthologous genomic sequence of the developmental *DACH* gene and its neighboring gene desert between human and mouse shows more than 1000 conserved regulatory sequences (Nobrega et al. 2003). Further comparison of the *DACH* sequence in evolutionarily distant vertebrates, including pufferfish, designates only 32 of them as deeply conserved (Nobrega et al. 2003). However, deletion of ultraconserved elements in mice leads to largely normal animal development, suggesting that deep conservation of enhancers does not always define crucial-for-survival enhancer function (Ahituv et al. 2007). In addition, sequence conservation alone does not hint necessarily to enhancer activity, as there are several non-enhancer regulatory sequences in the genome. But even if it corresponds to an enhancer, conservation does not provide any information about the spatiotemporal specific enhancer activity.

During the last years, identification of putative enhancer sequences has advanced due to the development of high-throughput sequencing techniques. One approach is the detection of accessible chromatin by DNase-seq (Crawford et al. 2005; Boyle et al. 2008; Hesselberth et al. 2009), FAIRE-seq (Giresi et al. 2007; Gaulton et al. 2010), MNase-seq (Yuan et al. 2005) or ATAC-seq (Buenrostro et al. 2013). Basic principle of these methods is that in regions that TFs bind, chromatin is depleted from nucleosomes and thus, accessible. However, these methods detect also other accessible regulatory regions in the genome, such as promoters, therefore requiring additional steps or complementary techniques to identify only the enhancer elements. To better specify the enhancer search, chromatin immunoprecipitation followed by massively parallel sequencing (ChIP-seq), either for TFs or for histone modifications is commonly used. Enhancers contain TFBSs that interact with coactivators such as the acetyltransferases P300/CBP (Eckner et al. 1994; Yao et al.

1998; Heintzman et al. 2007; Visel et al. 2009). *In vivo* mapping of P300 binding was found to be of very high prediction accuracy for enhancer regions, when the immunoprecipitated sequences were tested in mouse transgenic assays (Visel et al. 2009). Moreover, nucleosome depleted regions (NDRs) are flanked by regions marked by specific histone modifications, such as H3 lysine 4 monomethylation (H3K4me1) and H3 lysine 27 acetylation (H3K27ac) (Creyghton et al. 2010; Heintzman et al. 2009; Rada-Iglesias et al. 2011; Zentner et al. 2011). H3K4me1 is associated with active, poised and repressed enhancers, while H3K27ac is indicative of active enhancers. Further research revealed more classes of enhancers characterized by different combinations of histone modifications depending on their activity status (Ernst et al. 2011; Pradeepa et al. 2016). Finally, enhancers are often found in close proximity with the transcription machinery at core promoters. Thus, 3C-based methods, which assess the physical contact between DNA interacting partners, constitute another strategy to link enhancers with their target promoters (Section 1.2.1).

To test directly the function of identified regulatory sequences, transgenic reporter assays are employed. The different types of reporter assays, reviewed in (Kvon 2015), make use of the enhancer property to activate the transcription of a reporter gene from a minimal promoter. When integrated randomly in the genome, minimal promoter and reporter gene, i.e. *LacZ,* act as "enhancer traps". Their activation depends on the cumulative activities of surrounding regulatory elements (Kothary et al. 1989; Ruf et al. 2011). Following an "enhancer-reporter" strategy, the function of a selected enhancer is directly tested *in vivo* by cloning the potential element in front of a minimal promoter and a reporter gene. The integration can be either site-specific or not. A lot of model organisms have been used for transgenic enhancer reporter assays (*M. musculus*, *D. melanogaster*, *D. rerio* etc) and the transcription of the reporter gene can be assessed through RNA *in situ* hybridization (O'Kane and Gehring 1987), live imaging of the RNA transcript (Bothma et al. 2014), by performing LacZ staining (Kothary et al. 1989) or detecting GFP fluorescence (Chiocchetti et al. 1997). More advanced versions of these enhancer reporter assays are approaches of functional enhancer screenings, such as self-transcribing active regulatory region paired to sequencing (STARR-seq) (Arnold et al. 2013). STARR-seq tests candidate enhancer sequences by cloning them downstream of a TSS, so that active enhancers drive their own transcription. Due to simplified library construction, multiple enhancers were identified in flies and mammals using STARR-seq (Arnold et al. 2013; Muerdter et al. 2018).

Finally, manipulating regulatory sequences *in vivo* can lead to the ultimate validation of their function. The traditional gene targeting by homologous recombination was complemented and in many cases replaced by methods which achieve site-specific DNA cleavage using meganucleases

(Rouet et al. 1994), small zinc-finger proteins fused to endonucleases, the zinc-finger nucleases (ZFNs) (Bibikova et al. 2001, 2002, 2003) or transcription activator-like effector proteins fused to nucleases (TALENs) (Christian et al. 2010; Zhang et al. 2011; Reyon et al. 2012). Subsequently, genome editing has been revolutionized with the development of the CRISPR-Cas9 technology (Jinek et al. 2012; Cong et al. 2013; Mali et al. 2013; Jinek et al. 2013), which gave an unprecedented boost to the speed and complexity of testing CREs *in vivo*, among various genome targeting purposes.

1.1.3 CRISPR-Cas9, a powerful tool for *in vivo* functional testing of CREs

CRISPR-Cas9 is a molecular tool which derived from a prokaryotic defense mechanism and became a revolution in the field of genome editing. CRISPR stands for clustered regularly interspaced short palindromic repeats and it was first discovered in the genome of *Escherichia coli* (Ishino et al. 1987). Later, though, CRISPR loci were found in many archaea and bacteria (Mojica et al. 2000; Makarova et al. 2011).

In the following years, it was revealed that spacer sequences between CRISPR repeats are derived from plasmids and viruses (Bolotin et al. 2005; Mojica et al. 2005; Pourcel et al. 2005). Moreover, CRISPR loci are transcribed (crRNAs) and associated with the *Cas* gene (CRISPR-associated), which is adjacent to the repeat arrays and encodes for a protein with nuclease and helicase activity (Jansen et al. 2002; Haft et al. 2005). These findings suggested that the CRISPR loci are part of the immune system of microorganisms and act to keep memory of past infections and provide resistance to next infections by spacer-invading DNA sequence similarity (Makarova et al. 2006; Barrangou et al. 2007). The CRISPR system requires a single protein to identify and cleave the DNA (Jinek et al. 2012; Gasiunas et al. 2012) and a protospacer adjacent motif (PAM) next to the crRNA-targeted sequence of the invading DNA (Mojica et al. 2005; Shah et al. 2013). Moreover, the complementary strand to the CRISPR locus transcribes tracrRNAs (trans-activating crRNAs,) which together with the host endoribonuclease III and the Cas9 protein contribute to the production of mature crRNAs (Deltcheva et al. 2011). In 2012, Jinek et al. demonstrated that the tracrRNA:crRNA duplex directs the Cas9 to induce specific dsDNA cleavage (Jinek et al. 2012). Overall, these key findings summarize the three steps of this immune mechanism; adaptation, when invading DNA short sequences are inserted in the host CRISPR array; expression, transcription of mature crRNAs which constitute arrays of repeats between variable spacer sequences originating from the invading DNA; interference, as the foreign DNA sequence complementary to the crRNA spacer sequence gets cleaved by the Cas proteins.

The wide application of the CRISPR-Cas system in genome editing was better foreseen, since the dual tracrRNA:crRNA was engineered as a single guide RNA (sgRNA), consisting of two features: a 5' end 20nt sequence complementary to the target sequence and a 3' end double-stranded structure binding to Cas9 nuclease (Jinek et al. 2012). That way, any DNA sequence could be targeted by changing only the 5'end of the sgRNA as long as there is a PAM sequence next to it. As a result, in 2013, CRISPR-Cas9 was for first time used in human cancer cell lines and human pluripotent stem cells to edit the genome by co-expressing the different components of the system targeting either single or multiple loci (Cong et al. 2013; Jinek et al. 2013; Mali et al. 2013). These studies were only the beginning of the precision editing CRISPR era, when genes are engineered in numerous organisms, reviewed in (Sander and Joung 2014; Peng et al. 2014).

Upon Cas-induced cleavage, two cellular pathways, the Non-Homologous End Joining repair (NHEJ) and the Homology-Directed Repair (HDR), are recruited to repair the double strand breaks (DSBs). Cas9-induced DSBs have been used to introduce NHEJ-mediated insertions and deletions (indels) as well as to stimulate HDR with either double-stranded plasmid DNA or single-stranded oligonucleotide as donor templates. Using two sgRNAs to target two positions on a chromosome can also induce larger structural variations (SVs), e.g. deletions, inversions or duplications of the intermediate DNA fragment. Kraft and colleagues presented a fast protocol, named CRISVar, for engineering structural variants in mice (Kraft et al. 2015) (Figure 1. 1). In total, six different loci were targeted with pairs of sgRNAs in mouse embryonic stem cells (mESCs) and all sorts of structural variants, ranging from 1.1 kb to 1.6 Mb, were obtained.

Despite the extensive development and use of CRISPR, the main disadvantage of this technology remains the off-target effect. A series of studies have shown that Cas9 binds ectopically to other loci, introducing mutations (Fu et al. 2013; Hsu et al. 2013; Pattanayak et al. 2013). This depends on the number, position and distribution of mismatches throughout the entire sgRNA sequence. Furthermore, abundance of Cas9 is also a determining factor for off-target effects (Pattanayak et al. 2013). The more Cas9 is expressed, the more off-targeting occurs. However, altering the delivery method by transfecting the cells directly with the ribonucleoprotein complex sgRNA-Cas9 induces more transient Cas9 activity and less off-targets (Lin et al. 2014; Kim et al. 2014a). Although various studies have achieved to reduce partially the off-target effects of CRISPR by different approaches (Adli 2018), there is still progress to be made on improving the specificity of the system and decreasing the potential byproducts of this technology. Overall, CRISPR-Cas is one of the fastest evolving technologies and among other achievements, it has greatly contributed in understanding the role of nuclear chromatin folding in gene regulation.

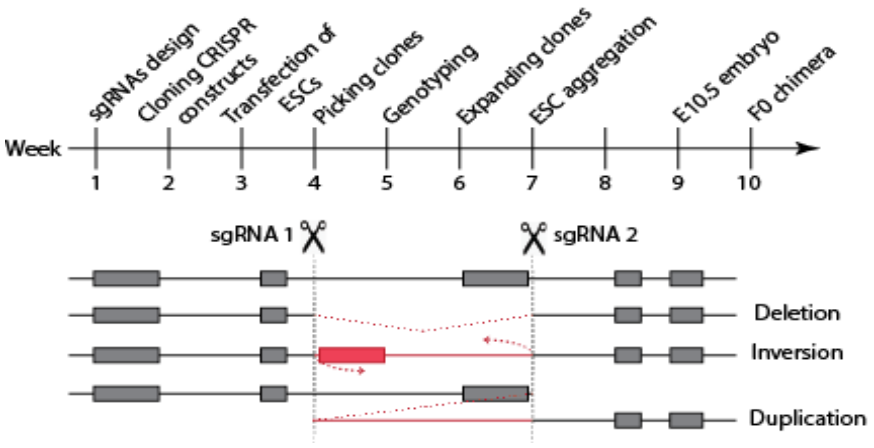


Figure 1. 1. Overview of the experimental setup to induce structural variations by CRISPR-Cas9

A pair of sgRNAs are designed to target a genomic sequence in mESCs. Using the protocol of Kraft et al., F0 transgenic mice derived from genetically edited mESCs can be born after 10 weeks. Using two sgRNAs can induce deletions, inversions and even duplications of the intervening targeted genomic sequence. Scheme is modified from (Kraft et al. 2015)(detailed protocol in Method section 3.1).

1.2 3D genome organization

To understand how gene regulation works, it is not only necessary to identify the position and function of regulatory elements. Instead, key answers are hidden in finding which enhancer interacts with which gene, how these -in many cases- long-range interactions are achieved in the nucleus, and what happens when they are disturbed. Along with high resolution microscopy techniques, e.g. fluorescence *in situ* hybridization (FISH), which have been a great tool to visualize nuclear organization (Boyle et al. 2001; Tanabe et al. 2002), the majority of recent discoveries are based on proximity ligation technology and its derivatives.

1.2.1 Conformation Chromosome Capture (3C)-based technologies

Basic principle of 3C-based assays is the fixation of native chromatin by formaldehyde crosslinking in order to capture the regions that are in close proximity in the nucleus. Then, the fixed chromatin is fragmented by restriction enzyme digestion and ligated, so that hybrid-molecules are generated. These 3C pairs are indicative of interacting partners in the 3D space of the nucleus and their abundance is used to produce an averaged contact frequency across the cell population in a particular state. 3C (one-to-one) was first used to capture the frequency of interactions between any two genomic loci in the *Saccharomyces cerevisiae* (yeast) (Dekker et al. 2002). Soon after, 3C applied to the beta-globin locus in mice showed that a cluster of enhancers, called the locus control region (LCR), comes in close proximity with the distally-located active genes only in the erythroid cells, looping out the inactive genes located between them (Tolhuis et al. 2002). As 3C followed by

semi-quantitative PCR or real-time PCR are quite laborious techniques, many variants were developed attempting to improve some aspects. The resolution, the sensitivity and the multiplexing ability of each of them constitute critical points that these various methods try to improve (Davies et al. 2017).

One of the 3C derivatives is circular chromosome conformation capture (4C) that identifies all potential interacting partners from one specific viewpoint in the genome (one-to-all) (Zhao et al. 2006; Simonis et al. 2006). The distinguishing feature is the inverse PCR with primers designed from the viewpoint to amplify all interacting partners in a 3C library. To achieve higher PCR efficiency, it is necessary that the circular fragments are small. Thus, an additional digestion with a second enzyme is performed followed by a second ligation (Simonis et al. 2006). To achieve library of high complexity, many PCRs need to be performed from a high number of cells. Advantages of the 4C are the good resolution and the high sensitivity, while one of the possible biases is how extensively the GC content of some sequences affects the PCR (Stadhouders et al. 2013). 4C has been successfully used for identifying promoter-enhancer interactions at specific loci (Andrey et al. 2013), for linking disease-associated SNPs with genes (Pasquali et al. 2014), and also for discovering structural variations, such as translocations (Simonis et al. 2009).

Alternatively, capture-C generates genome-wide interactions from hundreds of viewpoints (Hughes et al. 2014). In detail, 3C-libraries are sonicated and sequencing adapters with indexed barcodes are added. The library is then enriched for fragments of interest using biotinylated RNA capture probes designed for each viewpoint. Next, the captured fragments are pulled down with streptavidin beads, amplified and sequenced in a paired-end fashion. In this approach, due to the sonication step that produces unique DNA ends any potential duplicate fragments can be removed in the data analysis. Capture-C from ~400 viewpoints in limb-related genes has enabled the characterization of their respective regulatory landscapes in limb tissues and revealed two modes of chromatin folding: the tissue-specific and the tissue-invariant (Andrey et al. 2017). Another variant of this method, called capture Hi-C (cHi-C) generates all-to-all contacts, but custom-designed biotinylated probes capture only regions of interest. This approach is suitable for generating high resolution maps focused only the loci of interest.

The method of choice for detecting megabase-scale contacts and determining large chromatin structures is Hi-C. Hi-C generates contact maps between all parts of the genome (all-to-all). After digestion of the crosslinked chromatin, the fragments ends get filled in with biotinylated nucleotides and blunt ligation follows. Then, DNA is de-crosslinked, extracted and sonicated. Finally, all junction ligations are sequenced upon streptavidin pulldown. The resolution of this

approach has been greatly improved over the last few years from 40 kb to fragment size to single base-pair resolution, which makes it feasible to identify fine structures of the 3D genome (Lieberman-Aiden et al. 2009; Jin et al. 2013; Rao et al. 2014; Bonev et al. 2017).

3C technologies are steadily getting improved and more derivatives are being developed, so that biases get eliminated and resolution gets increased. Undoubtedly, due to the 3C-based methods, our understanding on enhancer-promoter communication and regulatory landscapes has been strengthened. It has also become clear that the 3D chromatin folding in the nucleus follows a hierarchical organization that enables the complex patterns of gene regulation during cell differentiation and development.

1.2.2 Organization of chromatin folding in the nucleus

Spatial organization of the chromatin in the nucleus is known to play important role in transcriptional regulation of genes in many organisms (Cremer and Cremer 2001; Sexton et al. 2007; Bickmore 2013). FISH experiments have demonstrated that chromatin is organized in chromosomal territories, which rarely intermix (Lichter et al. 1988; Pinkel et al. 1988). These findings were later confirmed using Hi-C for the first time in a human lymphoblastoid cell line (Lieberman-Aiden et al. 2009). A more recent study using Hi-C at the single cell level showed that 5-10% of each chromosome intermingles with other chromosomes (Stevens et al. 2017). Within the nuclear space, gene-rich chromosomes are mostly localized in the center, while gene-poor chromosomes are found in the nuclear periphery.

Based on the first Hi-C contact maps, it was established that chromosomes are separated in two types of domains, namely A and B compartments, which tend to interact with each other in a homotypic fashion (Lieberman-Aiden et al. 2009). In particular A compartment, which associates with transcriptionally active regions, interacts with other A compartments. Similarly, B compartments, which display features of repressed regions, interact with other B compartments. The physical separation of A and B compartments can also be observed by super-resolution microscopy and *in silico* modeling of single-cell Hi-C (Wang et al. 2016; Stevens et al. 2017; Nir et al. 2018). Higher resolution *in situ Hi-*C revealed that these two types can be further subdivided in six subcompartments in total (A1 & A2, B1, B2, B3 & B4) (Rao et al. 2014). The distinction between the different subcompartments is based on their histone modification status, gene positioning, transcriptional status, and replication timing. Thus, the genome at the super-megabase scale is organized according to its functional state and is thus segregated into physical nuclear regions with

different local concentration of transcriptional resources. Therefore, loci must switch compartments in order to be transcribed.

Improvements in the resolution of 3C-based experiments have further revealed the partitioning of the genome into domains of preferential chromatin interactions, the topologically associating domains (TADs) (Dixon et al. 2012; Nora et al. 2012; Sexton et al. 2012). TADs are self-interacting chromatin domains of up to ~3 Mb size and mostly stable between different cell types and across species (Dixon et al. 2012; Rao et al. 2014). Functionally, TADs direct enhancer-promoter interactions and insulate ectopic enhancer activities from neighboring TADs, thus shaping the regulatory landscape of genes. Insertion of regulatory sensors, consisting of a minimal promoter and a reporter gene, have been used to detect the regulatory information at a given locus (Kondo and Duboule 1999; Ruf et al. 2011; Marinić et al. 2013; Akhtar et al. 2013; Symmons et al. 2014, 2016). The extent of detected regulatory activity is in accordance with the extent of TADs. The reporter signal though disappears abruptly at the borders of the regulatory domains, which confirms what was observed also in Hi-C contact maps (Dixon et al. 2012). At these regions, known as TAD boundaries, the directionality of contacts changes and the insulation score between domains is increased. According to the computational method used and the resolution, the definition of a TAD and its boundaries varies (Andrey and Mundlos 2017). Domains with preferential interactions within them or else loop domains are frequently characterized by increased contact frequencies at the top corner, which represent loops formed between the boundaries of these domains (Rao et al. 2014). However, loops can also appear between regions of frequent contact within TADs, in a more tissue-specific fashion. These domains are known as sub-TADs and they are mostly associated with CTCF binding (Phillips-Cremins et al. 2013; Rao et al. 2014). Loops formed between boundaries have been further associated with architectural proteins binding, e.g. cohesin and CTCF, but also with active transcription marks e.g. H3K4me3, H3K36me3, housekeeping genes, and repeat elements (Dixon et al. 2012; Bonev et al. 2017). In *Drosophila*, while the importance of CTCF localization at TAD boundaries is under debate, active transcription seems to be critical for TAD formation (Hou et al. 2012; Ulianov et al. 2016; Hug et al. 2017; Chathoth and Zabet 2019). In mESCs, transcription participates in TAD insulation, but is not sufficient to form a TAD boundary (Bonev et al. 2017). While the functional significance of TADs and their boundaries in gene regulation is better understood, the factors underlying their formation are still to be investigated (Section 1.2.4).

TADs function as guides and their boundaries set the limits for enhancer-promoter interactions. Therefore, alterations in their structure or their boundaries can lead to ectopic enhancer-promoter communication followed by gene misexpression. Disruption of TAD structures due to structural

variations leading to enhancer adoption has been already reported in human patients (Lettice et al. 2011; Ibn-Salem et al. 2014; Lupiáñez et al. 2016). Similar structural variations can be induced in model organisms using genome editing methods, like CRISVar (Kraft et al. 2015). In mouse, deletion of the *Epha4* gene and its TAD boundary results in loss of insulation between the *Epha4* and the neighboring *Pax3* TADs and leads to *Pax3* misregulation in limbs by the enhancers of *Epha4*, thus causing a brachydactyly phenotype (Lupiáñez et al. 2015). Inversion of the centromeric boundary of the *Epha4* TAD brings the *Epha4* enhancers in the vicinity of the neighboring *Wnt6* gene inducing its ectopic expression in the limb buds and consequently, a condition called F-syndrome (Lupiáñez et al. 2015). Furthermore, genomic duplications including TAD boundaries can give rise to new insulated domains known as neo-TADs (Franke et al. 2016). When no gene is included, the regulatory information is restricted in the new domain preventing any gene misregulation. On the contrary, when a gene gets incorporated into the neo-TAD adopts its regulatory potential and can lead to disease (Franke et al. 2016). Accordingly, genomic duplications at the human *IGF2* locus including the gene and a lineage-specific superenhancer lead to *de novo* interaction of *IGF2* with the enhancer, thus causing *IGF2* overexpression in several colorectal cancer cases (Weischenfeldt et al. 2017). Overall, these studies demonstrate that TADs act to direct enhancer activities to their normal target genes, while preventing them from activating genes in other neighboring domains.

1.2.3 Dynamic *versus* stable enhancer-promoter communication

While TADs were described as rather stable regulatory units, there is a lot of variation in interactions observed at the sub-TAD level. These variations represent mostly tissue-specific enhancer-promoter loops which form unique intra-TAD structures. In particular, enhancers screen for their target promoter within the TAD they belong and transmit their activities to the appropriate promoter in a spatiotemporally-defined manner. Most vertebrate enhancers are mapped at ~20 to 50 kb away from their target promoters (Furlong and Levine 2018). However, there are exceptions such as the *ZRS* enhancer, which is located ~850 kb away from its target, the *Shh* promoter (Lettice et al. 2003). To achieve gene activation, it is necessary that the two interacting partners are in close proximity, even transiently. In some cases, proximity is sufficient to initiate gene activation and it even functions as a trigger. For instance, targeted tethering of a looping factor to the β-globin promoter induces forced looping with its LCR enhancer and ectopic transcriptional activity (Deng et al. 2012, 2014). By contrast, in other loci enhancers are already in proximity with their target promoters prior to gene expression (Spilianakis and Flavell 2004; Andrey et al. 2013; Ghavi-Helm et al. 2014; Cruz-Molina et al. 2017; Rubin et al. 2017a). These two modes of enhancer-promoter

communication during development and lineage commitment have been described as instructive and permissive, respectively (de Laat and Duboule 2013).

Most of the regulatory landscapes follow the instructive mode of interaction, so that chromatin looping to the target promoter takes place almost simultaneously with gene expression (Freire-Pritchett et al. 2017). These dynamic interactions act to refine transcriptional activities of enhancers. At the *SatB1* locus, the gene promoter establishes contacts with many regulatory elements in the flanking gene desert specifically in thymocytes where the gene is highly active (van de Werken et al. 2012). Conversely, in brain tissue where *SatB1* expression is significantly lower, 4C experiments did not detect any contact within the regulatory landscape. Furthermore, at the *Pitx1* locus, a dynamic chromatin conformation restricts the fore- and hindlimb *Pen* enhancer to act and induce *Pitx1* expression only in hindlimbs (Kragesteen et al. 2018). Interestingly, the instructive mode of enhancer-promoter communication has been associated with active or repressive chromatin marks, suggesting a mechanistic interplay between the two phenomena (Andrey et al. 2017).

While the instructive model seems to be interdependent from functional chromatin properties, the permissive model is less understood. Permissive or preformed enhancer-promoter contacts occur long before gene activation and are tissue-invariant. In *Drosophila*, it was reported that more than 90% of enhancer-promoter interactions are involved in stable loops independently of tissue and developmental stage and are associated with paused Polymerase (Ghavi-Helm et al. 2014). Similarly, pluripotent-associated enhancers are in contact with anterior neural gene promoters already in undifferentiated mESCs (Cruz-Molina et al. 2017). These poised enhancer-promoter associations are mediated by the polycomb repressive complex 2 (PRC2), which loss in differentiating mESCs severely compromises the induction of these anterior neural genes. Furthermore, architectural proteins like cohesin and CTCF, and other tissue-specific transcription factor binding have also been associated with pre-established enhancer contacts (Samstein et al. 2012; Eijkelenboom et al. 2013; Kim et al. 2014b; Rubin et al. 2017a; Andrey et al. 2017). Independently of the underlying mechanism, preformed loops are suggested to contribute to rapid and robust gene activation and transcription. Interestingly, at some loci pre-established contacts co-exist with dynamic ones (Montavon et al. 2011; Andrey et al. 2013), probably enabling different types of regulatory behaviors to act at the same time.

1.2.4 Shaping the 3D genome

Chromatin folding is a very dynamic process. It is established, maintained and potentially reset during cell cycle, differentiation and development. It is therefore essential to characterize the mechanisms shaping the 3D genome and thus, enabling long-range interactions. Over the past years, various factors have been associated with the genome architecture including the key protein complexes Mediator and cohesin as well as the CCCTC-binding factor (CTCF).

Mediator is a protein complex composed of 26 subunits in mammals. It was first discovered in yeast and shown to be involved in transcription activation (Kelleher et al. 1990; Flanagan et al. 1991). In particular, Mediator subunits, as targets of various TF activation domains, receive their regulatory signal and *transmit* it to the Pol II enzyme. Apart from Mediator's complex role in transcription, it has additionally been associated with chromatin looping mostly in combination with cohesin (Kagey et al. 2010). Cohesin is one of the structural maintenance of chromosomes (SMC) complexes and it was initially identified for its crucial role in holding together the sister chromatids from S-phase until mitosis, thus ensuring proper chromosome segregation (Michaelis et al. 1997; Guacci et al. 1997; Uhlmann and Nasmyth 1998). The ring-shaped structure of SMCs can function by surrounding the chromatin fibers (Haering et al. 2008). Interestingly, colocalization of Mediator and cohesin was suggested to mediate mostly short-range (<100kb) dynamic enhancer-promoter looping (Kagey et al. 2010; Phillips-Cremins et al. 2013).

In addition, cohesin is involved in linking distal elements located in Mb-scale distances, when colocalized with the DNA-binding protein, CTCF (Phillips-Cremins et al. 2013). CTCF is a highly conserved zinc-finger protein, known as the main insulator protein described in vertebrates. Based on heterologous reporter assays, CTCF was described as general transcription factor able to activate or repress gene expression (Baniahmad et al. 1990; Lobanenkov et al. 1990). However, later it was shown that the diverse biological functions of CTCF, reviewed in (Ong and Corces 2014), are probably due to its ability to mediate intra- and interchromosomal interactions as shown by 3C-based technologies at the *β-globin* and at the *Igf2/H19* gene loci (Ling et al. 2006; Splinter et al. 2006; Zhao et al. 2006). CTCF associates with various proteins in different genomic loci (Zlatanova and Caiafa 2009), but is required for cohesin recruitment to chromatin and for establishing long-range interactions (Parelho et al. 2008; Rubio et al. 2008; Wendt et al. 2008). Downregulation of cohesin by RNA interference led to disruption of CTCF-mediated intrachromosomal contacts and thus, further indicated the involvement of both architectural proteins in shaping long-range interactions (Hadjur et al. 2009). Moreover, analysis of looping events in 1% of the human genome revealed that 60% of them are tissue-specific and that the interacting regions are enriched in CTCF binding and

active histone marks (H3K4me1, me2, me3 and H3K27ac, H3K9ac) (Sanyal et al. 2012). Together these studies proposed that CTCF bridges tissue-specific enhancer-promoter communication (Phillips-Cremins et al. 2013; Vietri Rudan et al. 2015).

The partitioning of the genome in TADs revealed that CTCF is frequently found at TAD boundaries (Dixon et al. 2012) and acts to mediate stable chromatin interactions (Phillips-Cremins et al. 2013; Vietri Rudan et al. 2015; Andrey et al. 2017). Taking together the binding frequency of CTCF at boundaries and its involvement in stable looping interactions, it was hypothesized that anchor loops formed between TAD boundaries are CTCF-mediated. Indeed, Rao and co-workers showed, using Hi-C, that co-binding of CTCF and of the cohesin subunits, RAD21 and SMC3 are found at 80% of all observed loops (Rao et al. 2014). Additionally, 92% of the CTCF motif pairs were found to be in convergent orientation, i.e. towards one another (Rao et al. 2014). Inversion of both a genomic segment with few CTCF binding sites and of a single CTCF motif leads to disruption of chromatin structure and gene expression changes (de Wit et al. 2015; Guo et al. 2015). In an attempt to model all previous observations concerning CTCF/cohesin and chromatin folding, two research groups suggested loop extrusion (Nasmyth 2001) as the main mechanism behind chromatin domain formation (Sanborn et al. 2015; Fudenberg et al. 2016). The loop extrusion model describes an extrusion complex, e.g. cohesin, which entraps the DNA in its ring-shaped structure and forms a small loop. By moving along the chromatin, the loop gets bigger until it meets "roadblocks", which are CTCF sites in convergent orientation. This model explains, among other observations, the different experimental outcomes upon depletion of CTCF and cohesin. Initial efforts toward depletion of these proteins resulted in only mild effects on TAD organization probably due to incomplete loss or resolution problems (Seitan et al. 2013; Sofueva et al. 2013; Zuin et al. 2014). However, using the auxin-inducible degron system in mESCs, depletion of CTCF leads to loss of the anchor loops between TAD boundaries, but also decreased insulation between TADs (Nora et al. 2017). Similarly, degradation of cohesin subunits (RAD21 or SCC1) and deletion of NIPBL, the loader of cohesin onto chromatin, leads to significant loss of loop domains/TADs in all cases (Rao et al. 2017; Schwarzer et al. 2017; Wutz et al. 2017). Taken together, both cohesin and CTCF are essential for loop formation in accordance with the current loop extrusion model. Nevertheless, despite the significant changes observed upon CTCF or cohesin loss, A/B compartments were mostly left intact, which hints to a cohesin-independent mechanism of chromatin folding at this scale (Rao et al. 2017; Schwarzer et al. 2017). Finally, loss of TAD insulation, loops and TADs surprisingly caused only marginal gene expression changes (Nora et al. 2017; Rao et al. 2017; Schwarzer et al. 2017). Potentially, the timing needed for accumulation of transcriptional defects might be longer than

examined in the above studies. Moreover, investigation of such dynamic processes *in vitro* does not always constitute a representative model system to assess transcriptional effects and might thereby lead to misinterpretations. As such, the extent of significance of 3D chromatin folding on gene regulation during development needs to be further elucidated.

1.3 The developing limb as a model for gene regulation

Since decades the limb has been used as a paradigm of organ development, starting from undifferentiated cells which undergo complex patterning processes to ultimately form delicate structures. The signaling pathways and genes involved in limb development are largely conserved, but the resulting morphologies vary significantly between species. Developing limbs are easy to access, to observe and to handle experimentally and therefore constitute a powerful model to study gene regulation.

1.3.1 Limb development and molecular key players

All limbed tetrapods possess three different skeletal segments, known as stylopod, zeugopod and autopod (Figure 1. 2). These skeletal segments give rise to the humerus /femur, ulna and radius/tibia and fibula, carpals/tarsals and phalanges of the arms/legs, respectively. Although they are conserved among species, the number of skeletal parts within these segments and the proportions differ from each other. In fact, processes concerning early limb development are highly conserved among most species (Sears et al. 2015).

Limbs arise from the lateral plate mesoderm (LPM) of the flank initially as limb buds, consisting of non-differentiated mesenchymal cells covered by a layer of ectoderm. They develop along three major axes: the proximal-distal (PD) axis from the trunk of the body towards the hand, the dorsal-ventral (DV) axis from the back of the hand to the hand palm and the anterior-posterior (AP) axis from digit I to V (Figure 1. 2). Each axis is controlled by specific signaling centers and governed by key developmental genes, which proper coordination is crucial for limb development.

The PD axis outgrowth and patterning is under the control of the apical ectodermal ridge (AER). The AER is a layer of specialized epithelial cells that cover the limb bud tip and its formation is mainly driven by the family of fibroblast growth factors, *Fgfs*, expressed and secreted in the AER. Their main function is to keep the subjacent mesenchymal cells in a proliferative state. *Fgf8* is expressed from the appearance of the AER onwards and its expression is initially induced by FGF10, present in the LPM (Ohuchi et al. 1997). Additional members of the family are also expressed in the AER e.g. *Fgf4*, *Fgf9*, *Fgf17*, but individual and combinatorial knockout (KO) experiments have not resulted in any limb abnormalities (Mariani et al. 2008). On the contrary, inactivation of *Fgf8* resulted in limb

defects, supporting its critical role in normal limb development (Lewandoski et al. 2000). At the other edge of the axis, retinoic acid (RA) signaling originating from the flank has been suggested to initiate limb outgrowth and antagonize AER signaling (Mercader et al. 2000; Cooper et al. 2011; Rosello-Diez et al. 2011), but its exact role still needs to be determined.

At the AP axis, the main signaling center is a posterior-distal region of the limb bud mesenchyme known as zone of polarizing activity (ZPA). Grafting the posterior ZPA in the anterior side of wing buds led to mirror-image symmetrical digit duplications (Tickle 1981). The number and identity of duplicated digits was dependent on the concentration and duration of the signal, which hinted towards a morphogen (Tickle 1981). Later, it was found that sonic hedgehog (SHH) is the morphogen secreted by ZPA cells and that it defines all digits with the exception of digit I (Riddle et al. 1993). Proper AP axis development is based on patterning asymmetry determined by the solely posterior expression of *Shh*. However, the limb is asymmetrically pre-patterned before *Shh* is expressed due to the mutual antagonism between posterior HAND2 and anterior GLI3 factors (te Welscher et al. 2002). *Hand2* is positively regulated by all HOX9 factors (Xu and Wellik 2011) and RA signaling (Niederreither et al. 2002), while *Hoxa* and *Hoxd* gene expression is restricted posteriorly also by GLI3 (Zuniga and Zeller 1999; Kmita et al. 2005; Tarchini et al. 2006). All these factors together with FGF8 from the AER contribute to *Shh* expression initiation in the AER (Crossley et al. 1996; Capellini et al. 2006; Galli et al. 2010). Upon ZPA-SHH establishment, many of the prior activators of *Shh*, now become targets of SHH, as for example *Hoxd* genes (Tarchini and Duboule 2006). However, the main mediators of SHH signaling in the limb defining AP patterning are the glioma-associated oncogene proteins (GLI). In particular, SHH inhibits the cleavage of the GLI3 full-length protein into its truncated transcriptional repressor (GLI3R), creating a SHH-dependent gradient (Wang et al. 2000). Cells of digit V and IV are descendants of *Shh*-expressing cells of the ZPA and thus, see maximal SHH signaling due to autocrine signaling for different amount of time (Harfe et al. 2004). Digit III sees also maximal SHH but for shorter period of time and some of its cells are dependent on paracrine SHH (Harfe et al. 2004). Digit II is completely dependent on paracrine signaling and digit I is SHH independent, as it is the only remaining digit in *Shh* null animals (Harfe et al. 2004). Taken together, ZPA-SHH are primarily responsible for giving rise to the correct number and identity of the skeletal parts of the future hand and foot.

The DV axis specification originates from diverse epithelial-mesenchymal interactions (Chen and Johnson 1999, 2002). One of the main key players of this axis is *WNT7a* which is expressed from the dorsal ectoderm, while *EN-1* induced by BMP signaling is expressed from the ventral ectoderm. These two signals are antagonistic, as *En-1* KO experiments or impaired *Bmp* expression lead to

expansion of *Wnt7a* expression in the ventral ectoderm and distal limb structures appear with double dorsal characteristics (Loomis et al. 2003). Conversely, *Wnt7a*$^{-/-}$ *mice* have limbs with dorsal-to-ventral transformation (Parr and McMahon 1995). A downstream target of WNT7A is the LIM-homeodomain transcription factor *Lmx1b*, which is expressed in the dorsal mesenchyme, thus specifying the dorsal patterning of the limb (Riddle et al. 1995; Vogel et al. 1995).

For proper limb development the signaling feedback loops are interlinked. The AER is necessary for ZPA and *vice versa*. For example, BMP4 initially induces expression of its antagonist, *Grem1,* in the distal mesenchyme to create a functional AER (Nissim et al. 2006; Bénazet et al. 2009). Later, *Grem1* is responsive to SHH and transmits its signal to the AER-FGFs, which signals back to ZPA-SHH thus creating the SHH-GREM1-FGF epithelial-mesenchymal feedback loop (Zúñiga et al. 1999; Michos et al. 2004). This constitutes a self-regulatory limb signaling system which allows for proper development.

In summary, the limb is an organ which growth and patterning is controlled by interconnected signaling pathways and feedback loops. Therefore, it is challenging to identify individual roles of the included signaling molecules. Expression of all these molecules is thus necessary to be controlled not only spatially, but also temporally. Moreover, changes in any of these factors, pathways and spatiotemporal control lead to morphological alterations, morphological evolution, but also limb malformations.

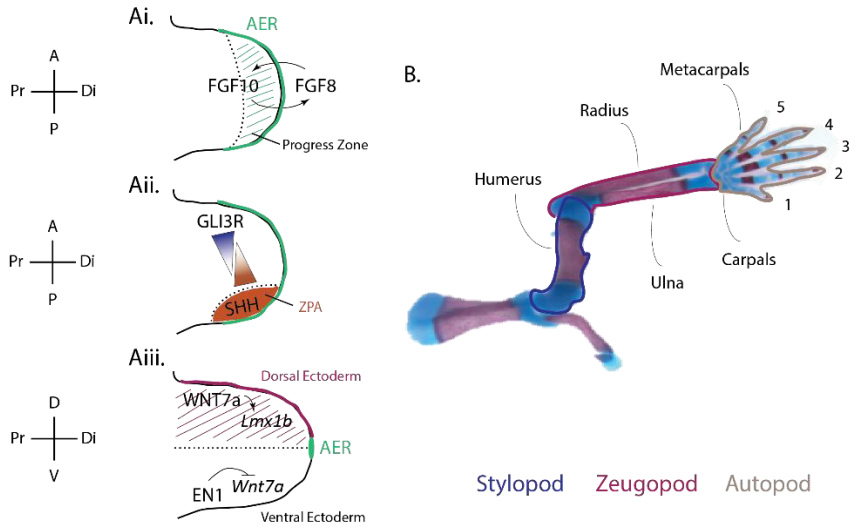


Figure 1. 2. Signaling pathways patterning the vertebrate limb bud and skeletal parts making a forelimb

Orientation of limb bud axes are indicated (A: anterior, P: posterior, Pr: proximal, Di: distal, D: dorsal, V: ventral). **A. i**, Proximal-distal outgrowth: FGF8 secreted from the AER (green perimetric line) creates a positive feedback loop with FGF10 in the progress zone (green-shaded part) that drives distal limb bud outgrowth. **A. ii**, Anterior-posterior patterning: ZPA (brown-shaded part), located in the posterior limb bud mesenchyme secretes SHH. The anterior-posterior gradient of SHH acts antagonistically with GLI3-R (dark

blue), repressor form of GLI3. **A. iii**, Dorso-ventral patterning: WNT7A, produced from the dorsal ectoderm (purple-shaded part) induces expression of *Lmx1b* in the underlying mesoderm which in turn controls dorsalization. EN-1 inhibits *Wnt7a* expression in the ventral ectoderm. Figure modified from Mundlos & Horn, 2014. **B.** Proximo-distal development of a forelimb consists of the stylopod giving rise to the humerus, zeugopod producing the radius and ulnas, and the autopod forming carpal and metacarpal bones and digits. The digit identity 1 to 5 is controlled from the anterio-posterior axis.

1.3.2 Sonic Hedgehog (*Shh*) gene and its regulation in the limb

One interesting observation is that independently of the underlying cause leading to limb morphological malformations or evolutionary adaptations, there are some commonalities. One relies on the fact that early developmental events, in the limb, are more conserved than later ones and that some pathways are more often found altered than others (Petit et al. 2017). For example, expression levels of SHH signaling, which specify distal skeletal elements of the limbs at late stages, vary more often than others (Petit et al. 2017). This could be either of biological significance or because it is one of the most studied signaling pathways during organ development.

Shh is a developmental gene which belongs to the Hedgehog family of genes. Hedgehog (*hh*) gene was first identified in *D. melanogaster* by genetic screens (Nüsslein-Volhard and Wieschaus 1980). In vertebrates, three genes are homologues of *hh*, Sonic Hedgehog (*Shh*), Indian Hedgehog (*Ihh*) and Desert Hedgehog (*Dhh*). All of them are morphogens that are secreted and diffuse across developing tissues thus creating concentration gradients necessary for normal body organization (Ingham and McMahon 2001). DHH is important in germ cell development in the testis and peripheral nerve sheath formation. IHH regulates bone and cartilage development and is partially redundant with SHH.

Taking a closer look at the *Shh* locus (Figure 1. 3), it comprises the *Shh* gene with three exons and a large gene desert upstream of its TSS. Almost all regulatory sequences defining the complex spatiotemporal expression pattern of *Shh* are found within this gene desert (Anderson et al. 2014). In particular, two intronic enhancers (SBE1, SFPE2) and one lying ~9 kb upstream of *Shh* (SFPE1) regulate its expression in the ventral midline of the spinal cord and hindbrain, the ventral midbrain and caudal region of diencephalon (Epstein et al. 1999). Three enhancers (SBE2, 3, 4) located 300-450 kb upstream of *Shh* control its forebrain expression (Jeong et al. 2006), while another three elements (MRCS1, MFCS4, MACS1) at a distance of 600-750 kb upstream of *Shh* dictate its transcription in the epithelial linings from the oral cavity to the gut along the anteroposterior axis (Sagai et al. 2009). More recently, two additional enhancers were reported; one (SLGE) driving *Shh* expression in the gut and lung of rodents (Tsukiji et al. 2014) and a second (SBE5) in the *zona limitans intrathalamica* (zli), a signaling center essential for forebrain development (Yao et al. 2016). The SLGE enhancer was also identified as a sequence regulating *Shh* in neural progenitor cells

(Benabdallah et al. 2016). The most distal regulatory sequence, located approximately 850 kb away from *Shh* TSS, is the ZPA regulatory sequence, *ZRS*, (also *MFSC1*) named after its activity in the ZPA of the limb (Lettice et al. 2003; Sagai et al. 2005). *Shh* KO experiments lead to defects in notochord and floor plate, skeletal malformations of the limbs and the ribs, and cyclopia (Chiang et al. 1996). This systemic failure of many developing organs underlines the role of SHH in development.

The limb enhancer of *Shh*, *ZRS*, was identified as a region frequently associated with congenital limb malformations, for instance preaxial polydactyly (Lettice et al. 2003). Trying to understand the molecular basis of this phenotype, Lettice and colleagues suggested the *ZRS* as the regulatory sequence responsible for initiation and spatial expression of *Shh* in the limb (Lettice et al. 2003). Homozygote deletion of the *ZRS* in mice resulted in complete loss of *Shh* expression and deformed zeugopod and autopod bearing only one digit (Sagai et al. 2005). Interestingly, the *ZRS* KO mice exhibit the same limb phenotype as the *Shh* KO mice, affirming the *ZRS* as the sole enhancer of *Shh* in the limb (Chiang et al. 2001). Indeed, both posterior-specific and temporal expression of *Shh* depend on transcriptional regulators interacting at the *ZRS*. HAND2 and posterior HOX factors are some of *Shh* upstream regulators initiating its expression in the limb bud (Kmita et al. 2005; Capellini et al. 2006; Tarchini et al. 2006; Galli et al. 2010). Chromatin immunoprecipitation experiments showed that HAND2 is bound at the *ZRS* (Galli et al. 2010) via E-box binding sites (Osterwalder et al. 2014). Similarly, several HOX binding sites have been mapped on the *ZRS* sequence and were suggested to control the temporal expression of *Shh* (Leal and Cohn 2016; Lettice et al. 2017). Three of them are important for initial binding of early expressed HOXD9-11 and establish the activity levels of *Shh*, whereas two sites at the 5' *ZRS* with a preference for HOXD13 binding maintain *Shh* levels at later stages of limb development (Lettice et al. 2017). Moreover, the spatial expression of *Shh* solely in the posterior limb bud is determined from a cluster of ETS transcription factor binding sites with opposing functions (Lettice et al. 2012). In particular, five ETS1/GABPα sites define the limit of the ZPA, whereas two ETV4/5 sites restrict the activity outside of the ZPA (Lettice et al. 2012). In summary, a number of activating and repressing binding sites at the *ZRS* compose a complex enhancer with a 5' part responsible for its spatiotemporal activity and a 3' part thought to be required for long-range activity (Lettice et al. 2017).

The *ZRS* enhancer's exceptionally high conservation among vertebrates (Gehrke and Shubin 2016) supports the functional importance of its regulatory role. As *Shh* expression and *ZRS* enhancer activity in the limb are identical, perturbations of the *ZRS* sequence itself or alteration in enhancer dosage can lead to *Shh* misregulation. In fact, more than twenty point mutations within the *ZRS* and many chromosomal duplications including the *ZRS* are reported in humans with limb

malformations, reviewed in (Anderson et al. 2012). These so-called "*ZRS*-associated syndromes" include limb phenotypes like preaxial polydactyly, syndactyly, triphalangeal thumb polysyndactyly and Werner mesomelic syndrome (WMS) (Wieczorek et al. 2010). The similarity of these malformations to mirror-image duplicated digits in chick graft experiments (Tickle 1981) drives the hypothesis that they are caused by *Shh* misexpression. Indeed, gain-of-function mutations or gain of copies of the *ZRS* lead to anterior ectopic expression of *Shh* in the limb and polydactyly phenotype observed in humans, mice, cats and chicken (Sharpe et al. 1999; Lettice et al. 2002, 2003, 2008; Dorshorst et al. 2010). For more severe human phenotypes, like WMS linked to long bone deficiency or Laurin-Sandrow syndrome associated with limb and nose defects, the actual mechanisms are not fully understood. However, for some mutations abolishment of repressive ETV sites or creation of new ETS sites underlies the ectopic activity (Lettice et al. 2012). Similarly, mutations of the *ZRS* were proposed to drive evolutionary changes. For instance, deletion of activator binding sites in pythons (ETS1, HOXD13) results in early arrest of *Shh* expression. Thus, developing snakes initially form vestigial hindlimbs, which regress within a few days of development (Kvon et al. 2016; Leal and Cohn 2016). Accordingly, progressive sequence degradation of the *ZRS* correlates with loss of activity in limbless snake species (Kvon et al. 2016). Despite the morphological plasticity of the limb as a developing organ, *Shh* and its unique limb *ZRS* enhancer display only little evolutionary variation. Although sequence conservation is important, efficient 3D enhancer-promoter communication is equally necessary to achieve robust spatiotemporal expression.

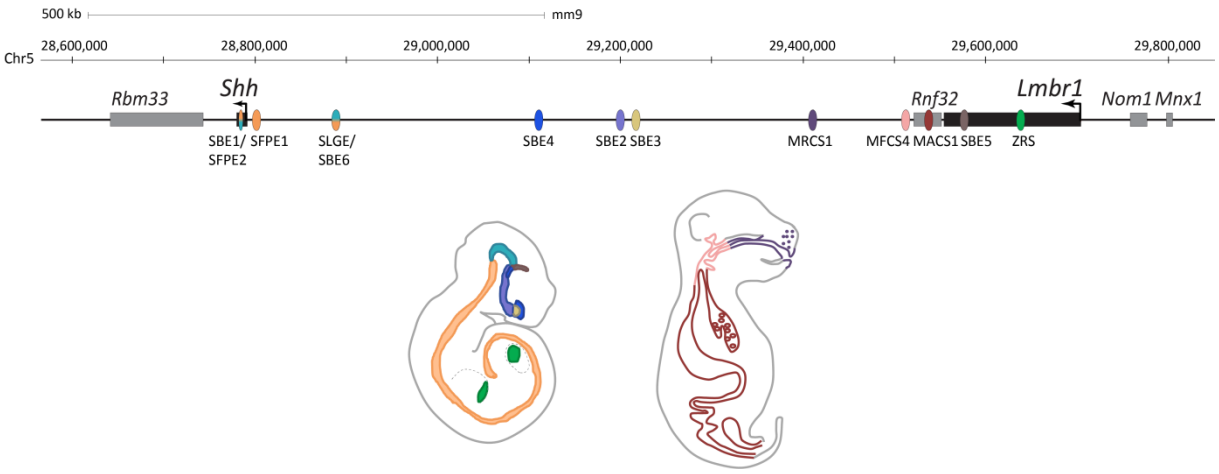


Figure 1. 3. Shh regulatory domain comprises numerous enhancers

The ~1 Mb genomic domain of the *Shh* gene with its regulatory elements which underlie its spatiotemporal pattern of expression. Coloured ovals representing *Shh* enhancers are scattered across the entire domain.

Below, schematic illustrations of an E11.5 embryo and *Shh*-expressing tissues color-coded according to the enhancer element controlling their regulation. *Shh* is expressed in the ventral midline of the spinal cord, hindbrain, the ventral midbrain and caudal region of diencephalon (SBE1, SFPE1, SFPE2), in forebrain (SBE2, 3, 4), in the *zona limitans intrathalamica* (zli) (SBE5), in the epithelial linings from the oral cavity to the gut along the anteroposterior axis (MRCS1, MFCS4, MACS1), in the gut and lung (SLGE) and in the limb (ZRS). Modified from (Anderson and Hill 2014).

1.3.3 3D Chromatin structure at the *Shh* locus

The gene regulation of *Shh* is controlled by many enhancers that are scattered in a large gene desert and within neighboring genes, each responsible for complex patterning of different developing organs. It is therefore important to understand whether the genomic position of each element is important and how distal enhancers come into close spatial proximity with the *Shh* promoter to generate these pleiotropic activities.

In an attempt to understand the well-coordinated *Shh* regulation, regulatory sensors were inserted and transposed in different positions along the genomic locus (Anderson et al. 2014; Symmons et al. 2016). All regulatory sensors at any position of the locus can recapitulate the expression pattern of *Shh* in all different tissues. However, *Shh* expression is not detected when sensors are inserted downstream of *Shh* and upstream of the housekeeping gene, *Lmbr1,* thus defining the limits of its regulatory domain. These limits coincide with the *Shh* TAD boundaries as shown by several studies (Dixon et al. 2012; Jin et al. 2013; Rao et al. 2014; Symmons et al. 2014). Within the telomeric boundary of the TAD the most distal enhancer *ZRS* is located, while *Shh* is found almost 1 Mb away at the opposite edge of the domain. Interestingly, while the other *Shh* enhancers induce similar levels of *LacZ* expression across the gene desert, the *ZRS* reporter levels are not equivalent across the domain (Anderson et al. 2014). In fact, the reporter expression in response to the *ZRS* activity is higher toward the TAD boundaries but lower within the gene desert, suggesting a *ZRS* regulatory mechanism independent of the linear genomic distance to the *Shh* gene promoter, but related to their proximity. Supporting these data, 4C-seq experiments in E11.5 posterior limb buds confirmed the constantly high *Shh-ZRS* interaction independently of the distance between them (Symmons et al. 2016). Moreover, 4C in anterior and medial limb, where *Shh* is not expressed, revealed similarly equal frequency of contact between the two elements. These observations have been confirmed by an orthogonal approach. 3D-FISH experiments in cell nuclei from posterior limb buds probing *Shh*, *ZRS* and the *SBE4* enhancer found in the middle of the locus showed that *Shh-ZRS* distance is shorter than either *Shh-SBE4* or *ZRS-SBE4*, suggesting an interaction model where *Shh* and *ZRS* are in close proximity and the intervening gene desert is looped out (Amano et al. 2009; Williamson et al. 2016). Same finding was described for the anterior limb bud, the proximal limb and the flank at all time points examined, independently of *Shh* expression (Amano et al. 2009; Williamson et al.

2016). However, when *Shh* is transcriptionally active in ZPA, the colocalization between *Shh* and the *ZRS* increases (Williamson et al. 2016). Furthermore, Symmons et al. tested the relation of the *Shh-ZRS* linear distance and of an intact TAD configuration in promoting their 3D proximity (Symmons et al. 2016). Genomic inversions encompassing several *Shh* enhancers, including the *ZRS*, disrupt the TAD structure. As a consequence, enhancers' interactions with *Shh* promoter are interrupted and redirected toward the neighboring TAD, resulting in severe systemic phenotypes in mouse embryos. However, the bigger the inversion, the closer the *ZRS* is brought to the *Shh* gene. Interestingly, this reduced linear distance between *Shh* and the *ZRS* overcomes the perturbed TAD configuration, thus enabling the development of few digits.

Overall, these studies propose that the linear genomic distance between *Shh* and its most distal regulator, the *ZRS*, differs from their topological distance in a cell nucleus. The *Shh-ZRS* long-range regulation is possible due to their 3D proximity, which is achieved through the chromatin folding of the locus. Yet, some questions remain to be investigated; what is the role of this close proximity prior and independently of gene transcription and which are the underlying factors supporting it.

2 AIM OF THE PROJECT

The body patterning is orchestrated by highly defined gene expression patterns. Such patterns are thought to be controlled by the integration of signaling pathways at the level of gene regulatory landscapes. Specifically, gene transcription is regulated by –usually– multiple enhancers with redundant or additive activities. These enhancers come in close proximity to their target gene promoters in the nucleus and activate gene transcription in a spatiotemporal manner. Nevertheless, in some loci, enhancers are found close to promoters prior to gene activation.

SHH is a morphogen responsible for normal development of many organs, among which, the limbs. Despite that *Shh* genomic locus is populated by several enhancers, only one, the *ZRS*, regulates *Shh* in the limbs. Deletion of the *ZRS* in mice leads to digit aplasia, resembling the *Shh* KO phenotype. Although *Shh* and *ZRS* are located ~1 Mb away from each other, they are in proximity in the nucleus not only in the posterior *Shh*-expressing cells of the limb, but also in the anterior non-expressing cells suggesting a tissue-invariant interaction. Yet, the driving mechanism underlying this permissive chromatin structure and its effect on *Shh* regulation are not known.

The present work studies *Shh* locus *in vivo* as a paradigm of tissue-invariant enhancer-promoter interaction in the nucleus. First, to identify factors associated with *Shh* regulatory landscape and to detect changes between tissues, ATAC-seq and ChIP-seq are performed in mESCs, in midbrain and in the limb. Moreover, 3D chromatin interactions established in the three tissues are analyzed by using 4C-seq and Capture Hi-C. Second, to test the role of the 3D architecture on *Shh* regulation, key players associated with it are genetically perturbed by applying CRISPR-Cas9 in mESCs. The transcriptional effect is further assessed by expression analysis in E10.5 embryonic limb buds and phenotypic evaluation experiments, while the potential 3D changes at the locus are determined by producing Capture Hi-C maps from the same tissue. Overall, the aim of this study is to understand *Shh* regulation in the developing limb and whether it is modulated by the preformed 3D architecture of the locus.

3 MATERIALS & METHODS

All standard molecular biological procedures were performed according to (Sambrook and Russell 2001).

3.1 CRISPR-Cas9 genome editing

All deletion alleles of this study were generated by using the CRISPR-Cas9 editing system according to (Kraft et al. 2015) and (Andrey and Spielmann 2017) .

3.1.1 Cloning of gRNAs in the PX459 vector

The pSpCas9(BB)-2A-Puro (PX459) vector was obtained from Addgene (Section 9.1). Single guide RNAs (sgRNAs) were designed using the CRISPR Design Tool by the Zhang lab (http://www.genome-engineering.org/crispr/). The Benchling website was used as an alternative tool (https://benchling.com/). Both websites provide quality and off-target scores. The selected oligonucleotides had a quality score over 80% and less than two exonic off-target regions, when possible. All sgRNAs are listed in Table 9. 1. The sgRNAs were synthetized and cloned into the PX459 vector according to (Andrey and Spielmann 2017). Colonies were subject to PCR according to the following conditions:

Reagents	Volume (µl)
DNA	
10X Taq Buffer	2
dNTPs (12mM)	0.1
gRNA FP	0.1
Col RP	0.1
Taq polymerase	0.1
Bid H2O	
Total	20

Temperature	Duration	Cycles
94°C	4 min	
94°C	30 sec	
58°C	30 sec	x 30
72°C	30 sec	
72°C	5 min	
4°C	∞	

Table 3. 1 Colony PCR reaction

Positive colonies were grown for Midi-Prep and the DNA was extracted using the Nucleobond Xtra Midi (#740410), according to manufacturer's instructions. In order to verify the integration of the gRNA sequence in the PX459 vector, Sanger sequencing of the plasmid DNA followed. All Sanger sequencing reactions were performed using the custom-designed sequencing primer ColR (5'-CACGCGCTAAAAACGGA-3') and the BigDye v3.1 kit (Applied Biosystems), according to manufacturer's instructions. The samples were sequenced using an ABI3700 capillary sequencer

(Applied Biosystems) located at the Institute of Medical and Human Genetics at the Charité - Universitätsmedizin of Berlin.

Reagents	Volume (µl)
Plasmid	x (50-200 ng)
5X Seq Buffer	1
Big Dye v3.1	0.5
Primer	1
Bid H2O	y
Total	5

Temperature	Duration	Cycles
96°C	1 min	
96°C	30 sec	
50°C	30 sec	x 30
60°C	4 min	
4°C	∞	

Table 3. 2 Sequencing reaction for plasmids

3.1.2 Culturing CD1 / DR4 feeders cells

Feeder cells support the mouse embryonic stem cell (mESC) culture by providing them with nutrients preventing their differentiation. Primary embryonic fibroblasts (EMFIs), used in our laboratory, originated from E13.5 to E14.5 CD1 and DR4 mouse embryos. The feeders were cultured in regular Dulbecco's Modified Eagle's Medium (DMEM) 4,500 mg/ ml glucose without sodium pyruvate (Lonza, #BE12-733F), complemented with 10% regular fetal calf serum (FCS Superior, Biochrom, #S0615), 1x glutamine (100x, Lonza, #BE17-605E) and 1x penicillin / streptomycin (100x, Lonza, #DE17-603). After having tested for Mycoplasma contamination using the Mycoalert detection kit (Lonza, #LT07-118) and the Mycoalert assay control set (Lonza, #LT07-518), the cells got expanded and treated with Mitomycin C (Sigma, #M-4287 or M-0503) for mitotic inactivation. Then, they were frozen in cryovials at a density of 2.5 x 10^6 cells/ vial in regular feeder medium containing 20 % FCS and 20 % DMSO (Sigma, #D-2650).

3.1.3 Culturing G4 embryonic stem cells (ESCs)

G4 ES cells (129/Sv x C57BL/6 F1 hybrid strain) were seeded on plates or wells, which were first coated with 0.1 % gelatin (Sigma, #G-1393) and covered with a layer of feeder cells. The coating of the plates was complete after 30 min of incubation with gelatin at 37° C. Then, the remaining gelatin was removed and feeder cells were plated in a density of 3-4 x 10^4/ cm^2. After at least 6 hours, ES cells were seeded on top of the feeder layer in ESC medium, consisting of Dulbecco's Modified Eagle's Medium (DMEM) 4,500 mg/ ml glucose with sodium pyruvate (Gibco, #10829-018), complemented with 15 % regular fetal calf serum (FCS Superior, Biochrom, #S0615), 1x glutamine (100x, Lonza, #BE17-605E) and 1x penicillin / streptomycin (100x, Lonza, #DE17-603), 1x non-essential aminoacids (Gibco, #11140-35), 1x fresh 10 mM β-mercaproethanol (2-ME, Sigma #M-7522), 1x nucleosides (Chemicon #ES-008D) and 0.01% of LIF (Murine Leukemia Inhibitory Factor ESGRO (10^7

U/ ml, Chemicon, #ESG1107)). The ES+LIF medium needed to be changed every 24 hours. ES cells were frozen at a density of 1 x 10^6 cells/ vial in medium consisting of one-to-one regular ESC medium complemented with 5% FCS (final 20% FCS) and 2x ES cell freezing medium (DMEM 4.500 mg/ ml glucose with sodium pyruvate, 20% FCS, 20% DMSO (Sigma #D-2650)).

3.1.4 Transfection

To target the ES cells with the CRISPR guides (see Section 3.1.1), 3 x 10^5 G4 cells were seeded per well of a 6-well plate in ES+LIF medium. The wells were already coated with gelatin and a CD1 feeder layer. After ON incubation, the medium was changed to ES+LIF medium without penicillin /streptomycin. Then, 8µg of PX459 vector containing one gRNA was transfected using the FuGENE HD transfection reagent (Promega #TM-328) following manufacturer's instructions. The medium was changed to ES+LIF with Pen/Strep after approximately 12 hours and cells were left to recover for 24 hours. Then, the ESCs were split onto 6 cm dishes coated with gelatin and DR4 resistant-to-puromycin feeders and selection with puromycin followed for 2 days (final concentration 2µg/ ml, Sigma #P8833). Next, the selection media was replaced for ES+LIF to allow cells to recover and grow as clones.

3.1.5 Picking positive clones

Clones, that were successfully transfected, survived the puromycin selection and could be picked after 4-6 days of recovery. The dishes were washed with D-PBS (Dulbeccos's Phosphate Buffer Saline, Gibco™, Thermo Fischer Scientific, #14190169) and the picked clones were transferred into 96-well-plates in individual wells that already contained Trypsin (1x Trypsin-EDTA 0.5g/ l, Gibco #25300-054). After 10 min incubation at 37° C, ES+LIF was used to stop the trypsin reaction. The clones were left to grow for 2-3 more days in the 96-well plate. ES+LIF medium was changed every day.

3.1.6 Splitting and Freezing clones

After 2-3 days, the clones in the 96-well plate were ready to be split in three parts and 2/3 were frozen in 96-U-shaped-well plates with short term freezing medium, while the rest was expanded for DNA preparation and further genotyping. The short term freezing medium consists of one-to-one F1 and F2 media. Both media contain Bicarbonate Free DMEM (Gibco #52100) /10 mM Hepes (Sigma #H-0887) supplemented with 15% FCS, 1x glutamine 200mM (100x, Lonza #BE17-605E), 1x penicillin/streptomycin (100x, Lonza, #DE17-603), 1x non-essential aminoacids (100x, Gibco, #11140-35), 1x fresh 10 mM β-mercaproethanol (2-ME, 100x, Sigma #M-7522) and 1x nucleosides (100x, Chemicon #ES-008D). F1 consists of Bicarbonate free DMEM / 10mM Hepes with 20% FCS, while F2

is Bicarbonate free DMEM / 10mM Hepes supplemented with 20% FCS and 20% DMSO (Sigma #D-2650). To genotype the clones, they were first washed with D-PBS (Gibco™ #14190169) and lysed ON at 60° C in 50 µl of Lysis buffer (10 mM TrisCl pH 7.5, 10 mM EDTA pH 8.0, 10 mM NaCl, 0.5% sacrosyl and fresh 1:100 Proteinase K (final concentration 1 mg/ml)). The DNA was extracted using the KingFischer Flex Purification System (Thermo Scientific, #5400630). 50µl of lysate were mixed with 32 µl of 85% EtOH and 9 µl of the MagAttract Suspension G beads (Qiagen, #1026883) in each well of the 96-well plate. Two more plates with 100 µl of 85% EtOH (washing plates) and one with 35 µl of Bid H_2O (elution plate) were prepared. All the plates were loaded in the machine according to manufacturer's instructions. Initially, the samples get very well mixed and DNA binds to the magnetic surface of the beads. The DNA-bead mixture gets washed two times with 85% EtOH and dried in RT for around 10 min. Finally, DNA gets eluted in 35 µl of Bid H_2O and the beads get discarded.

3.1.7 Genotyping clones

3.1.7.1 PCR

To genotype the ESC clones, standard PCRs were performed according to (Sambrook and Russell 2001). Typing buffer 10x (Typ) is custom-made (200 mM Tris pH 8.8, 100 mM NH_4SO_4, 100 mM KCl, 20 mM MgSO4 and 1% Triton-X 100) and Taq polymerase is produced in-house by A.C. Stiege. Genotyping primers were designed with Primer3 (Table 9. 2). When a single gRNA was used, then the primers were designed at approximately 50 bp up- and downstream of the targeted deletion, so that the small deletions can be easily detected. When two gRNAs were used, then then primers spanned regions more than 100 bp up- and downstream of the deletion. The size of the band was visualized on 1% agarose gel. However, the CTCF deletions were less than 100 bp and the difference between the two alleles was visible only in 3-4% agarose gels.

Reagents	Volume (µl)
Typ buffer 10x	2
dNTPs (10 mM)	0.5
FP (10 mM)	0.75
RP (10 mM)	0.75
Taq polymerase	0.5
Total	20

Temperature	Duration	Cycles
94°C	4 min	
94°C	30 sec	
58-60°C	30 sec	x 5
72°C	2 min	
94°C	30 sec	
58-60°C	30 sec	x 30
72°C	1 min	
72°C	7 min	
4°C	∞	

Table 3. 3 Genotyping PCR

3.1.7.2 Quantitative Real-Time PCR (qRT-PCR)

To confirm that the CTCF motif was completely deleted in both alleles, copy number analysis was performed by qRT-PCR using the GoTaq qPCR Master Mix (Promega) on ABIPrism 7900 HT thermocycler. The primers were custom designed, so that one of them spans the CTCF motif. Additional set of primers spanning outside of the targeted region were designed to validate that the deletions were not exceeding more than the expected length. Approximately 5 ng of genomic DNA were used in each reaction and wildtype (WT) DNA was used as control. All qPCR primers for genotyping are listed in Table 9. 3.

3.1.7.3 Sub-cloning TA-GFP

To genotype both alleles of the selected clones, the pTA-GFP vector (Section 9.1) was used. Upon PCR amplification, the genomic DNA carried A-tails due to the Taq polymerase. In addition, the TA-vector carries T-tails, which allows for ligation of the complementary overhangs. The ligation reaction was performed in RT for at least 15 min (optimally 2h) as follows:

Reagents	Volume (µl)
pTA vector	-
PCR product	2
10x T4 Ligation buffer	2
PEG4000	2
ATP (10 mM)	2
Ligase	1
Bid H_2O	11
Total	20

Table 3. 4 Ligation reaction

The pTA-clone DNA was transformed into Top10 bacteria (in-house by A.C. Stiege) , according to standard protocols (Sambrook and Russell 2001). After approximately 12-16h of incubation at 37° C, the white colonies were picked and grown for mini preparation and subsequent sequencing. Mini-preps were performed using the Nucleospin® Plasmid EasyPure kit (Macherey-Nagel, #740727).

3.1.7.4 Sanger sequencing

All PCR reactions for Sanger sequencing were performed with 10-100ng DNA as a template and the BigDye v3.1 kit (Applied Biosystems), according to manufacturer's instructions. The samples were sequenced using an ABI3700 capillary sequencer (Applied Biosystems) located at the Institute of Medical and Human Genetics at the Charité - Universitätsmedizin of Berlin.

Reagents	Volume (µl)
PCR product	x (10-100 ng)
5x Seq Buffer	1.75 µl
BigDye v3.1	0.25 µl
Primer (10 mM)	0.5
Bid H_2O	Y
Total	5

Temperature	Duration	Cycles
96°C	1 min	
96°C	30 sec	
50°C	30 sec	x 25
60°C	4 min	
4- 8°C	∞	

Table 3. 5 Sequencing reaction for PCR products

3.1.8 Thawing and expanding selected clones

As soon as positive ones were selected, they were expanded. One of the freezing plates was thawed shortly at 37° C and the selected clones were transferred in 15 ml falcon tubes with 4 ml of ESC medium to dilute the DMSO of the F2 freezing medium. The cells were pelleted and seeded on gelatin-coated and feeder-covered 96-well plates in ES+LIF medium. The clones were gradually transferred in bigger culture plates until reaching good density in the 6-well plates, when they were split and frozen in cryovials (Section 3.1.6). The ESC+LIF medium was changed every day. A part of the cells was lysed as described in Section 3.1.6 and used for DNA extraction and genotyping.

3.1.8.1 Genotyping expanded clones

After expansion, the clones are to a great extent depleted from feeder cells and the genotyping is more reliable. Therefore, the clone DNA was genotyped with all the methods described before (Section 3.1.7) and one clone was selected for aggregation.

3.1.9 Generation and genotyping of transgenic mice

The selected ESC clones were used for generating mouse embryos and live animals by aggregating them with di- or tetraploid mouse embryos (Artus and Hadjantonakis 2011). The complementation experiments were performed by Lars Witler and Carol Macura at the Transgenics Facility (MPIMG). The molecular experiments were performed on mice either at embryonic stage E10.5 (~ 34-5 somites) or at E18.5. Genotyping of mouse embryos was performed on DNA extracted from the amnion or the tail tip. Genotyping of born mice was performed on DNA extracted from ear biopsies. DNA was extracted by adding 50 µl of QuickExtract (QuickExtract DNA Extraction Solution, Biozym #QE09050) and incubating for 20 min at 65° C. The reaction was quenched by increasing the temperature at 98° C for 2 min.

3.2 Expression analysis

Forelimbs and hindlimbs at stage E10.5 were dissected in 1x PBS/ DEPC and pooled together. Each biological replicate consisted of two forelimbs and two hindlimbs from the same individual. PBS was then removed carefully and the tissue was snap frozen in liquid N_2. The expression analysis was performed using three different methods: quantitative RT-PCR (qRT-PCR), RNA-seq and Whole-Mount In Situ Hybridization (WISH).

3.2.1 RNA extraction

Tissues were thawed on ice and the RNA extraction was performed using the RNeasy Mini kit (Qiagen #74106). Buffer RLT supplemented with β-mercaptoethanol was added onto the samples, according to manufacturer's instructions. Homogenization of the tissue was achieved by filtering the samples using a 0.5 mm syringe and careful pipetting, until cell clumps were dissolved. Then, equal volume of 100% EtOH was added onto the samples and the RNA was extracted based on the manufacturer's guidelines. Final elution was in 50 µl and the concentration was measured by Nanodrop.

3.2.2 cDNA

Samples were reverse transcribed using the SuperScript™ II First-Strand Synthesis System (Invitrogen #11904018) according to manufacturer's instructions:

Reagents	Volume (µl)
10 pg – 5 µg RNA	X
Random Hexamer Primers	1
dNTPs (10 mM)	1
Bid H2O	up to 13
65° C	**5 min**
5x First Strand buffer	4
0.1 M DTT	1
RNAse OUT	1
SS II	1
Total	20

Temperature	Duration
25° C	10 min
42° C	50 min
70° C	15 min
4° C	∞

Table 3. 6 cDNA reaction

After the reaction, the newly transcribed cDNAs were diluted to 5 ng/ µl and stored in -20° C freezer.

3.2.3 Quantitative Real-Time PCR (qRT-PCR)

In order to quantify the expression changes of the mutants of this study, we performed qRT-PCR using the GoTaq® qPCR Master Mix (Promega) on ABIPrism 7900 HT thermocycler. Primers (Shh FP: 5'- ACCCCGACATCATATTTAAGGA-3', Shh RP: 5'- TTAACTTGTCTTTGCACCTCTGA- 3', Rps9 FP: 5'- GACCAGGAGCTAAAGTTGATTGGA -3', Rps9 RP: 5' – TCTTGGCCAGGGTAAACTTGA- 3') were designed with the online tool Primer 3. 5 ng of cDNA were used per reaction and the *Rps9* gene was used as housekeeping gene. The primer sequences are:

The Master Mix contains already a proprietary dsDNA-binding dye, a low level of carboxy-X-rhodamine (CXR) reference dye, GoTaq Hot Start Polymerase, $MgCl_2$, dNTPs and a proprietary reaction buffer. Additional CXR was added to the mix before use. The conditions were:

Reagents	Volume (µl)
cDNA (5 ng /µl)	1
2X GoTaq Master Mix	10
Mix FP-RP (3 µM each)	2
Bid H_2O	7
Total	20

Temperature	Duration	Cycles
50° C	2 min	
95° C	10 min	
95° C	15 sec	X 40
60° C	1 min	
95° C	15 sec	
60° C	15 sec	
95° C	15 sec	

Table 3. 7 qRT-PCR reaction

To quantify the average abundance of the mRNA, 3 – 6 biological replicates (2 fore- and 2 hindlimb buds / biological replicate) and 3 technical replicates per sample were used. The fold change between wildtype and mutant samples was calculated using the delta delta Ct method (ΔΔCt) (Livak and Schmittgen 2001). For statistical analysis, one-sided Student's t-test was used. P-value was defined as follows: p*<0.05, p**<0.01 and p***<0.001.

3.2.4 RNAseq

Forelimb and hindlimb buds of 34-5 somite stage embryos (E10.5) were microdissected in cold PBS, snap frozen in liquid N_2 and immediately stored at -80°C. To isolate the RNA, tissues were thawed on ice and 500µl of TRIzol were added. Homogenization of the tissue was achieved by filtering the samples using a 0.4 mm syringe, until cell clumps were dissolved. Then, 200µl of Chloroform were added and the samples were mixed vigorously and centrifuged at 12000 xg for 15 min at 4°C. Then, the upper phase was transferred to new Eppendorf tubes and 500 µl of Isopropanol were added. After 10 min incubation at RT, another centrifugation step at maximum speed for 10 min at 4°C

followed. Pellet was washed two times with 100% and 70% EtOH accordingly, centrifuged, air-dried for 10 min and eluted with Nuclease-Free water.

3.2.4.1 Sequencing and analysis

RNA-seq samples were polyA enriched and sequenced using Illumina technology and following standard protocols at the Sequencing Core Facility of the Max-Planck-Institute for Molecular Genetics. Two biological replicates were used for each experiment.

Analysis of the samples was performed based on the pipeline established by Stefan Haas (Department of Computational Molecular Biology, MPIMG). Specifically, 50 bp paired-end reads were mapped to the mouse reference genome (mm9) using the STAR mapper version 2.4.2a with default settings besides the following options: outFilterMultimapNmax=5; outFilterMismatchNoverLmax=0.1; alignIntronMin=20; alignIntronMax=500000; chimSegmentMin=10. Reads per gene were counted based on the UCSC annotation tracks 'known genes' and 'RefSeq' combined via shared exon boundaries. The counting was implemented by applying the R function 'summarizeOverlaps' with 'mode=Union' and 'fragments=TRUE'. Finally, differential expression analysis was performed with the DEseq2.

3.2.5 Whole - mount *In Situ* Hybridization (WISH)

All solutions used for whole-mount *in situ* hybridization are listed in Table 9. 4. Buffers and solutions were treated with DEPC to inactivate RNase enzymes. WISH protocol was established and supervised by Norbert Brieske (Development and Disease Group, MPIMG).

3.2.5.1 Dissection and Fixation - Day I and II

Embryos were dissected in 1x PBS and fixed in 4 % PFA/ PBS at 4°C overnight. Fixed embryos were washed twice with PBST and dehydrated in increasing serial Methanol dilutions in PBST (25 %, 50 %, 75 % Methanol/ PBST, 2x 100 % Methanol, 10 min in each dilution) and stored at -20°C.

3.2.5.2 Hybridization – Day III

Prior to hybridization, embryos were rehydrated in 75 %, 50 % and 25 % Methanol/ PBST and washed twice with PBST (each step for 10 min). Then, embryos were bleached in 6 % hydrogen peroxide/ PBST for 1 hour on ice and washed in PBST. According to the embryonic stage, the embryos were further treated with 20 µg/ml Proteinase K (E10.5 for 2 min, E12.5 for 5 min) and the reaction was quenched by adding 2x with PBST/ glycine. Then, embryos were washed 5x with PBST and fixed for 20 min in 4 % PFA/ 0.2 % glutaraldehyde in PBS/ 0.1 % Tween 20 at RT. The fixation was stopped by washing 5x in PBST (5 min each) at 4°C. Then, embryos were washed 1x for 10 min in PBST/ L1 (1:1)

buffer at 68°C, incubated for 5 min in L1 buffer at 68° C and then for 1-2 h in H1 buffer (Table 9. 4). The DIG probes were prepared with standard lab protocols and stored in -80°C. Before use, they were diluted in H1 buffer and denatured for 10 min in 80°C. After cooling down, the probes were added onto the wells with the embryos for ON hybridization at 68°C.

3.2.5.3 Removing unbound DIG-labelled probe – Day IV

All buffers (L1, L2, L3) were heated to 68°C. The embryos were washed sequentially 3x 30 min in L1, 3x 30 min in L2 and 1x 15 min in L3 buffer always at 68°C. After cooling down to room temperature, the embryos were washed in 1:1 L3 buffer/ RNase solution for 5 min. They were then incubated 2x 30 min in RNase solution containing 100 µg/ ml RNaseA at 37°C and moved to 1:1 RNase solution/TBST-1 for 5 min. After 3x 5 min washing in TBST-1 at RT, the embryos were incubated in blocking solution for 2h shaking at RT. The antibody anti-Digoxygenin-AP (Roche, # 11093274910) was also diluted in blocking solution (1:5000) and incubated for 1h rotating at 4°C. Finally, antibody solution was added to the embryos and ON incubation on a shaker followed.

3.2.5.4 Removing unbound Antibody – Day V

Unbound antibody was removed by 5x 5 min washing with TBST-2, supplemented with 0.05% levamisole/ tetramisole at RT. Additional washing (8x 30-60 min) with TBST-2/0.05% levamisole/ tetramisole and ON incubation at 4°C in the same solution followed.

3.2.5.5 Staining – Day VI

Embryos were washed 3x 20 min in alkaline phosphatase buffer and antibody detection was carried out in BM Purple AP-substrate (Roche, #1442074) until a clear signal appeared. Embryos were then washed twice in alkaline phosphatase buffer, fixed in 4 % PFA/ PBS/ 0,2 % glutaraldehyde and 5mM EDTA and stored at 4°C. The stained limb buds were imaged using Zeiss Discovery V.12 microscope and Leica DFC420 digital camera.

3.3 Skeletal preparation

Embryos at E18.5 were sacrificed by induced hypothermia and were frozen at -20°C. To remove the skin, the embryos were warmed up in tap water at 65°C for 1 min. The embryos were then eviscerated, transferred in glass vials and fixed ON in 100% technical EtOH in RT. To stain the cartilage of the embryos blue, EtOH was replaced by Alcian Blue (150 mg/ l Alcian Blue 8GX (Sigma-Aldrich, #A5268)) and stained ON in RT. Upon second fixation of the embryos in 100% technical EtOH ON, they were treated with 1% KOH for 15 min for some partial tissue digestion. Then, the membranous bone was stained red using Alizarin Red solution (50 mg/l Alizarin Red (Sigma-Aldrich, # A5533) in 0.2 % KOH/ bid H_2O). Staining was performed for up to 2 days with visual inspection of

each specimen for proper red staining. Remaining tissue was gradually digested in 0.2% KOH-25% glycerol solution. Digestion was completely quenched by placing preparations in 25% glycerol for further clearing and 30% glycerol for short-term storage. Documentation of the skeletal preparations was done in either 25% or 30% glycerol. For long-term storage, 60% glycerol was used. The stained embryos were imaged using Zeiss Discovery V.12 microscope and Leica DFC420 digital camera.

3.4 Assay for Transposase - Accessible Chromatin with high throughput sequencing (ATAC-seq)

ATAC-seq is used for mapping chromatin accessibility genome-wide. The experiments were performed according to (Buenrostro et al. 2015) and conditions were optimized for the given tissue. Primary E10.5 limb buds were dissected and washed once in 1x PBS. The tissues were homogenized using the Ultra Turrax T8 disperser (IKA). Then, 50000 cells / biological replicate were washed in cold D-PBS (Gibco™ #14190169) and lysed in fresh lysis buffer (10mM TrisCl pH7.4, 10mM NaCl, 3mM $MgCl_2$, 0.1% (v/v) Igepal CA-630) for 10 min while being centrifuged. Next, the supernatant was discarded and the cells were prepared for the transposition reaction using the Nextera Tn5 Transposase (Nextera kit, Illumina #FC-121-1030). After 30 min at 37° C, the solution containing the nuclei was purified using the MinElute PCR Purification kit (Qiagen, #28004), the transposed DNA was eluted in 10µl of Elution buffer and stored in -20° C, if not immediately used. Barcoded adapters were added to the transposed fragments by PCR using the NEBNext® High Fidelity 2x PCR Master Mix (NEB #M0541).

Reagents	Volume (µl)
PCR-amplified DNA	5
NEB High Fidelity 2x MM	5
Nextera PCR Primer 1 (25 µM)	0.25
Nextera PCR Primer 2 (25 µM)	0.25
100x SYBR Green I	0.09
Bid H2O	4.41
Total	20

Temperature	Duration	Cycles
98° C	30 sec	
98° C	10 sec	
63° C	30 sec	X 20
72° C	1 min	
95° C	15 sec	
60° C	15 sec	
95° C	15 sec	

Table 3. 8 ATAC-seq PCR reaction

To avoid saturation after 5 cycles of PCR an aliquot (5 µl) was used to perform qPCR so that the needed number of cycles are defined.

Reagents	Volume (µl)	Temperature	Duration	Cycles

Transposed DNA	10	72°C	5 min	
NEB High Fidelity 2x MM	25	98°C	30 sec	
Primer 1 (25 µM)	2.5	98°C	10 sec	
Barcoded Primer 2 (25 µM)	2.5	63°C	30 sec	x 5
Bid H_2O	10	72°C	1 min	
Total	50	4°C	∞	

Table 3. 9 ATAC-seq test qPCR

The remaining 45 µl of the PCR reaction were amplified for the additional number of cycles. The total number was never more than 12. Finally, the samples were purified using the AMPure XP beads (Agencourt, #A63881) and eluted in 20 µl. Concentration was measured using Nanodrop and Qbit and the quality of the samples was estimated by the Sequencing Core Facility at MPIMG using Bioanalyzer 2100 (Agilent). Next, ATAC-seq samples were sequenced 50 or 75 bp paired-end and each experiment was performed in duplicates. ATAC-seq pipeline was established in collaboration with Johannes Helmuth (Department of Computational Molecular Biology, MPIMG) and data analyzed as in (Paliou et al. 2019).

Sequence Name	Sequence
ATAC_Ad_1	AATGATACGGCGACCACCGAGATCTACACTCGTCGGCAGCGTCAGATGTG
ATAC_Ad_2.1	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGGAGATGT
ATAC_Ad_2.2	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGGAGATGT
ATAC_Ad_2.3	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGGAGATGT
ATAC_Ad_2.4	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGGAGATGT
Nextera qPCR Primer 1	5'-AATGATACGGCGACCACCGA
Nextera qPCR Primer 2	5'-CAAGCAGAAGACGGCATACGA

Table 3. 10. ATAC-seq sequencing adaptors and qPCR primers

3.5 Tissue collection and fixation for ChIP-seq, 4C-seq and Capture Hi-C

Limbs from E10.5 embryos were microdissected in 1% PBS and pooled together. Limbs were washed once with 1% PBS and then homogenized in 500µl Collagenase solution (0,1% Collagenase type 1a (Sigma #C9891), 0,1% (w/v) Trypsin, 5% FCS or Chicken Serum in DMEM:HAM'S F-12 (1:1)) for approximately 15 min in a Thermomixer. Additional disruption of cell clumps was achieved by using a 0.4 mm needle. Then, samples were transferred in a 50 ml falcon tube through a 40 µm cell strainer and complemented with 10% FCS/PBS. Samples were retransferred in a 15 ml falcon tube

before fixation. Formaldehyde 37% diluted to a final concentration of 1% for ChIP experiments and 2% for 4C-seq and Capture Hi-C was used to fix the samples (10min, on a rocker, RT). To quench the fixation, 1 mL of 1.425 M Glycine was added. Formaldehyde solution was removed by centrifugation (300 xg, 8 min) and fresh lysis buffer (10mM Tris pH 7.5, 10mM NaCl, 5mM $MgCl_2$, 0.1mM EGTA complemented with Protease Inhibitor) was added to isolate the nuclei. The samples were incubated for 10min on ice, centrifuged for 5min at 480 xg, washed with 1x PBS and snap frozen in liquid N_2.

3.6 Chromatin Immunoprecipitation paired sequencing (ChIP-seq)

Chromatin from at least 16 pairs of E10.5 limb buds were used for each ChIP experiment. ChIP-seq for CTCF and Rad21 was performed using the iDeal Kit for Transcription Factors (Diagenode, # C01010055), according to manufacturer's instructions. For histone modifications, the immunoprecipitation was performed as described previously (Lee et al. 2006; Jerković et al. 2017). 20 µg of chromatin were used for TF ChIP, while 10-15 µg were used for histone ChIP. The antibodies used in this study were: H3K27ac (Diagenode, #C15410174), H3K36me3 (Abcam, #ab9050), CTCF (Diagenode, #C15410210) and RAD21 (Abcam, #ab992). The TF ChIP-seq experiments were performed in duplicates (CTCF for *Wildtype*, *ΔCTCF i4i5*, *ΔCTCF i4i5ZRS*) or in singletons. The histone ChIP-seqs were performed in duplicates.

3.6.1.1 Sonication of chromatin

Samples were thawed on ice and 300 µl of the sonication buffer from the iDeal kit were added. After 10 min incubation on ice, the samples were then sonicated in a size range of 200-500bp using the Bioruptor UCD-300 (Diagenode) as following: 30´´ ON- 30´´ OFF for 50 cycles. Every 10 min the samples were vortexed and spinned shortly. Samples could then be stored at -80 °C and a 10 µl aliquot was saved for testing the sonication efficiency. The aliquot was reverse-crosslinked ON and next day the sonified chromatin was precipitated and concentration was measured by Nanodrop. Then, it was tested on a 1% gel for the sonication profile. For the histone ChIP, the conditions were identical, but the buffers were prepared according to (Lee et al. 2006; Jerković et al. 2017).

3.6.1.2 Immunoprecipitation

The amount of chromatin for each ChIP was calculated based on the aliquot concentration. According to iDeal Kit manual, the beads were blocked with 5% BSA and then incubated with the antibody rotating for at least 4h at 4° C. 1 µg of CTCF (Diagenode, #C15410210) and 8µg of Rad21 (Abcam, #ab992) antibody per sample were used. 20 µg of chromatin was then added and the mixture was incubated ON at 4° C. To remove any unspecific binding the mixture was washed

several times with a series of wash buffers provided by the manufacturer. Then, the immunoprecipitated DNA was eluted and incubated ON at 65° C for reverse-crosslinking. The decrosslinked chromatin was then precipitated, purified and eluted in 25 µl. For histone ChIP, the samples were processed according to (Lee et al. 2006; Jerković et al. 2017).

3.6.1.3 Sequencing and analysis

Libraries were prepared using the Nextera adaptors and were sequenced single-end 50 or 75bp reads at the Sequencing Core Facility (MPIMG). Reads were mapped to the reference NCBI37/mm9 genome using Bowtie-2.2.6, then filtered for mapping quality ≥10 and duplicates were removed using Picard. Reads were extended (chromatin modifications: 300 bp, transcription factors: 200 bp) and scaled (one million / total of unique reads) to produce coverage tracks. For figure display purposes, some replicate ChIP-seq BigWig files were merged. BigWig files were visualized in the UCSC browser. The pipeline was established by Robert Schöpflin and further optimized by Guillaume Andrey and Daniel Ibrahim (Development and Disease group, MPIMG).

3.7 3C technologies

3.7.1 4C-seq

The 4C-seq libraries were performed as described previously (van de Werken et al. 2012; Franke et al. 2016). In summary, 4-bp cutters (DpnII, NEB, #R0543S and Csp6I, Thermo Fischer, #ER0211) were used as primary and secondary restriction enzymes. For each viewpoint, a total of 1 to 1.6 µg DNA was amplified by PCR. with the following primers associated with the respective restriction enzymes: Shh_read primer: 5'-CTACACGACGCTCTTCCGATCTCCATCGCAGCCCCAGTCT-3', Shh_reverse primer: 5'-CAGACGTGTGCTCTTCCGATCTCCATCCCCAGATGTGAGTGT-3'. All samples were sequenced 100 bp paired-end with Illumina Hi-Seq technology according to standard protocols at the Sequencing Core Facility (MPIMG). 4C-seq experiments from all viewpoints were carried out in duplicates.

3.7.1.1 Sequencing and analysis

Biological replicates were merged on the raw read level. Reads were filtered for the primer sequence, including the first restriction enzyme DpnII. After preprocessing, clipped reads were mapped to the reference NCBI37/mm9 genome, using BWA-MEM (v0.7.12-r1044) (Li and Durbin 2009) with default settings. Establishment of the pipeline, pre-processing, filtering and mapping of the samples were performed by Verena Heinrich (Department of Computational Molecular Biology, MPIMG). 4C-seq contacts were then analyzed in the murine region chr5:28000000-30000000. To calculate read count profiles the viewpoint and adjacent fragments 2 kb up- and downstream were

removed. A sliding window of 5 fragments was chosen to smooth the data and data was normalized to reads per million mapped reads (RPM). To compare interaction profiles of different samples, either subtraction or the log2 fold change for each window of normalized reads was obtained.

3.7.2 Capture Hi-C and Virtual Capture-C

Re-ligated products were then sheared using a Covaris sonicator (duty cycle: 10%, intensity: 5, cycles per burst: 200, time: 2 cycles of 60 s each, set mode: frequency sweeping, temperature: 4 to 7 °C). Adaptors were added to the sheared DNA and amplified according to Agilent instructions for Illumina sequencing. The library was hybridized to the custom-designed SureSelect beads and indexed for sequencing (75 bp paired-end) following Agilent instructions. Sonication, library preparation and sequencing was performed at the Sequencing Core facility (MPIMG). The cHi-C SureSelect library was designed over the genomic interval (mm9, chr5:27800001-30600000) using the SureDesign tool from Agilent. Capture Hi-C experiments were performed as singletons. As an internal control, we compared the results from six experiments for regions outside of the region of interest (chr16:91,000,000-91,550,000 and chr17:26,340,001-27,200,000).

3.7.2.1 Sequencing and analysis

CHi-C processing was performed by Robert Schöpflin (Development and Disease group, MPIMG). Raw sequencing reads were preprocessed with cutadapt v1.15 (Martin 2011) to trim potential low quality bases (-q 20 -m 25) and Illumina sequencing adapters (-a and -A option with Illumina TruSeq adapter sequences according to the cutadapt documentation) at the 3' end of reads. Next, sequencing reads were mapped to reference genome mm9, filtered and deduplicated using the HiCUP pipeline v6.1.0 (Wingett et al. 2015) (no size selection, Nofill: 1, Format: Sanger). The pipeline was set up with Bowtie2 v2.3.4.1 (Langmead and Salzberg 2012) for short read mapping. In case replicates were available, they were combined after the processing with the HiCUP pipeline. Juicer command line tools v1.7.6 (Durand et al. 2016) was used to generate binned contact maps from valid and unique reads pairs with MAPQ≥30 and to normalize maps by Knights and Ruiz matrix balancing (Knight and Ruiz 2013; Rao et al. 2014; Durand et al. 2016). For binning and normalization, only the genomic region chr5:27,800,001- 30,600,000 enriched in the DNA-capturing step was considered. Therefore, only reads pairs mapping to this region were kept, shifted by the offset of 27,800,000 bp and imported with Juicer tools using a custom chrom.sizes file which contained only the length of the enriched region (2.8 Mb). Afterwards, KR normalized maps were exported for 5kb and 10kb resolution and coordinates were shifted back to their original values. Subtraction maps were generated from KR normalized maps, which were normalized in a pair-wise manner before subtraction. To account for differences between two maps in their distance-dependent signal

decay, maps were scaled jointly across their sub-diagonals. Therefore, the values of each sub-diagonal of one map were divided by the sum of this sub-diagonal and multiplied by the average of these sums from both maps. Afterwards, the maps were scaled by 10^6 / total sum. CHi-C maps of count values, as well as subtraction maps, were visualized as heatmaps in which values above the 97-th percentile were truncated for visualization purposes.

In order to obtain more fine-grained interaction profiles for selected loci, we defined 10kb viewpoint regions and generated virtual Capture-C-like profiles based on the filtered, unique read-pairs that were also used for the cHi-C maps. A read pair was considered in a profile, when it had a MAPQ≥30 and one read mapped to the defined viewpoint region, while the other read mapped outside of it. The reads outside of the viewpoint were counted per restriction fragment and read counts were binned afterwards to 1 kb bins. In case a fragment was overlapping more than one bin, the read count was distributed proportionally. Afterwards, each binned profile was smoothed by averaging over a running window of 5 bins and scaled by 10^3 / sum of all its counts in the enriched region on chr5. The viewpoint and a window ± 5kb around it were not considered for the computation of the scaling factor. The profiles were generated with custom Java code using htsjdk v2.12.0 (https://samtools.github.io/htsjdk/).

Virtual Capture-C viewpoints: Proximal CTCF site (chr5:28,777,001-28,787,000), *Shh* promoter (chr5:28,789,001-28,799,000), *ZRS* enhancer (chr5:29,637,001-29,647,000).

3.8 3D polymer modeling

To investigate the spatial conformations of the *Shh* chromatin region we employed the *Strings & Binders Switch* (SBS) model (Barbieri et al. 2012; Chiariello et al. 2016), in collaboration with the Nicodemi lab in Naples, Italy. To estimate the minimal SBS polymer model that best reproduces the folding of the *Shh* region, i.e. the best distribution of the different binding sites along the polymer chain, the PRISMR method was used (Bianco et al. 2018). The genomic region that was modelled (chr5: 27800001-30600000) (mm9) encompasses the mouse *Shh* gene in the limb wildtype and *ΔCTCF i4:i5* cases based on the limb cHi-C interactions data (10 kb resolution). All details about the PRISMR method and MD simulations are described in (Bianco et al. 2018; Chiariello et al. 2016). The physical distances were investigated among the regions of interest. More precisely, the changes in relative distance among *Shh* and its regulatory regions were computed as $(d_{WT} - d_{i4i5})/d_{WT}$, d_{WT} and d_{i4i5} being the average distances among the highlighted region in limb wildtype and *ΔCTCF i4:i5*, respectively. The distribution of distances between *Shh* and its enhancer *ZRS*, normalized by their average distance in the limb wildtype, is statistically different in the limb wildtype (red) and *ΔCTCF i4:i5* (blue) cases (p-value < 10^{-3}, Mann-Whitney test). All simulations and computational analyses

were performed by Andrea Esposito, Andrea M. Chiariello, Simona Bianco and Carlo Annunziatella under the supervision of Mario Nicodemi (Dipartimento di Fisica, Università di Napoli Federico II, and INFN Napoli, Complesso Universitario di Monte Sant'Angelo, Naples, Italy).

4 RESULTS

4.1 3D chromatin structure and *cis*-regulatory landscape at the *Shh* locus

The *Shh* locus includes many CREs controlling *Shh* expression in various tissues (Figure 1. 3). Many of these elements have complementary and overlapping activities. To characterize the spatiotemporal expression of *Shh*, we performed WISH in wildtype E10.5 mouse embryos (Figure 4. 1). At this developmental stage, *Shh* is strongly expressed in the distal posterior part of the limb, but also in the ventral midbrain and in the *zona limitans intrathalamica* (zli) of the mouse brain. Then, comparing *Shh* expression with the *ZRS* activity (Amano et al. 2017), we observed that the *ZRS* enhancer drives *LacZ* expression only in the exact same domain of the limb (Figure 4. 1). Therefore, the *Shh* locus represents a curious case, where an enhancer activity coincides perfectly with the expression of the gene it regulates. It is thus intriguing to understand how *Shh* - *ZRS* communication is achieved in the 3D space of the nucleus, so that proper limb development is ensured. In order to answer this question, the regulatory landscape and 3D chromatin structure of the *Shh* locus were characterized in different tissues and cell types.

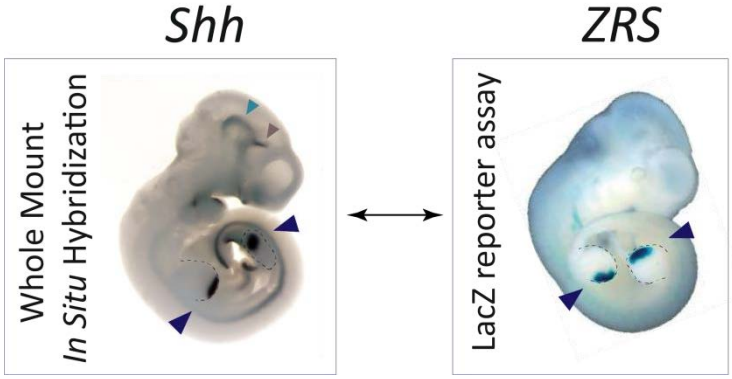


Figure 4. 1. Comparison of Shh expression with ZRS enhancer activity in E10.5 mouse embryos.

In the left box, *Shh* expression appears in the distal posterior part of the E10.5 developing limbs assessed by WISH (blue arrowheads). Light blue arrowhead depicts *Shh* expression in ventral midbrain (MB) and brown arrowhead in *zona limitans intrathalamica* (zli) of the mouse brain. In the right box, *LacZ* reporter assay shows the *ZRS* activity restricted in the same domain of the limbs (blue arrowheads), adapted from (Amano et al. 2017).

4.1.1 3D architecture of the *Shh* locus in E10.5 mouse limb buds

To assess how the *Shh* regulatory domain is folded in the nucleus, capture Hi-C (cHi-C) experiments were performed in somite-staged E10.5 wildtype limb buds, when *Shh* is highly expressed (Figure 4. 1). In detail, cHi-C probes were designed to cover an extended *Shh* region of 2.8 Mb (chr5: 27,800,001-30,600,000, mm9), thereby enriching chromatin interactions occurring at this region.

Over this genomic interval, interaction domains were visible, including the 1 Mb long *Shh* TAD (Figure 4. 2).

The *Shh* TAD comprises the *Shh*, *Rnf32*, *Lmbr1* genes and a large gene desert separating the two latter genes from *Shh* (Figure 4. 2A). *Shh* resides at the centromeric edge of the TAD and the *ZRS* enhancer ~850 kb away, within the intron 5 of *Lmbr1* gene at the telomeric edge. Aside from chromatin domains, specific interactions are visible from the cHi-C map. The most frequent interactions at the *Shh* locus are seen at the top corner of the *Shh* TAD (Figure 4. 2 A, black box) and have been previously described as chromatin loops (Rao 2014). In fact, these loops occur between the *Shh* region and the region around the *ZRS*. To better visualize this finding, virtual capture-C profiles (vC) were produced based on the cHi-C data, using the *Shh* promoter as viewpoint (Figure 4. 2 B). These profiles display clear interaction peaks between *Shh* and the region around the *ZRS,* but also with the *ZRS* itself.

To further examine the regulatory landscape of the locus, we used ATAC-seq and ChIP-seq on E10.5 embryonic limb buds (Figure 4. 2 C, D). The chromatin was accessible at the *Shh* promoter and at the *ZRS* enhancer as well as at the promoters of the two genes, *Lmbr1* and *Rnf32,* which are expressed in numerous tissues. In addition, using ChIP-seq, we observed enrichment of the H3K27ac mark, associated with active enhancers, at the *ZRS* enhancer. In the same condition, we could not detect H3K27ac-enrichment throughout the gene desert, thus suggesting the absence of other active limb enhancers.

As *Shh* and *ZRS* are both located within the boundaries of the *Shh* TAD, we assessed whether we could detect local enrichment of the boundary-associated protein CTCF using ChIP-seq in E10.5 limb buds (Figure 4. 2E). Two binding events were detected on each side of *Shh* and three (i4, i5, i9) around the *ZRS*, named after the number of intron they are embedded in. We then scanned the captured region for CTCF motif enrichment using the online tool FIMO (Find Individual Motif Occurrences (http://meme-suite.org/tools/fimo) and the CTCF motif from the JASPAR Core database (Figure 4. 7A). At each binding site, we could identify a CTCF motif and its orientation. Interestingly, the centromeric CTCF binding sites have the same orientation, which is convergent to the i4, i5 and i9 CTCF orientation. This observation is in agreement with the current model that TAD formation is mediated by loop extrusion between convergent CTCF binding sites (Sanborn et al. 2015; Fudenberg et al. 2016).

Accordingly, we concluded that both *Shh* and the *ZRS* reside within TAD boundaries and are surrounded by CTCF binding sites. As TAD boundaries are suggested to form stable chromatin loops, this configuration likely poises the *Shh-ZRS* interaction in the nucleus. It is however unknown how the *Shh* locus is organized in the 3D nuclear space in other tissues.

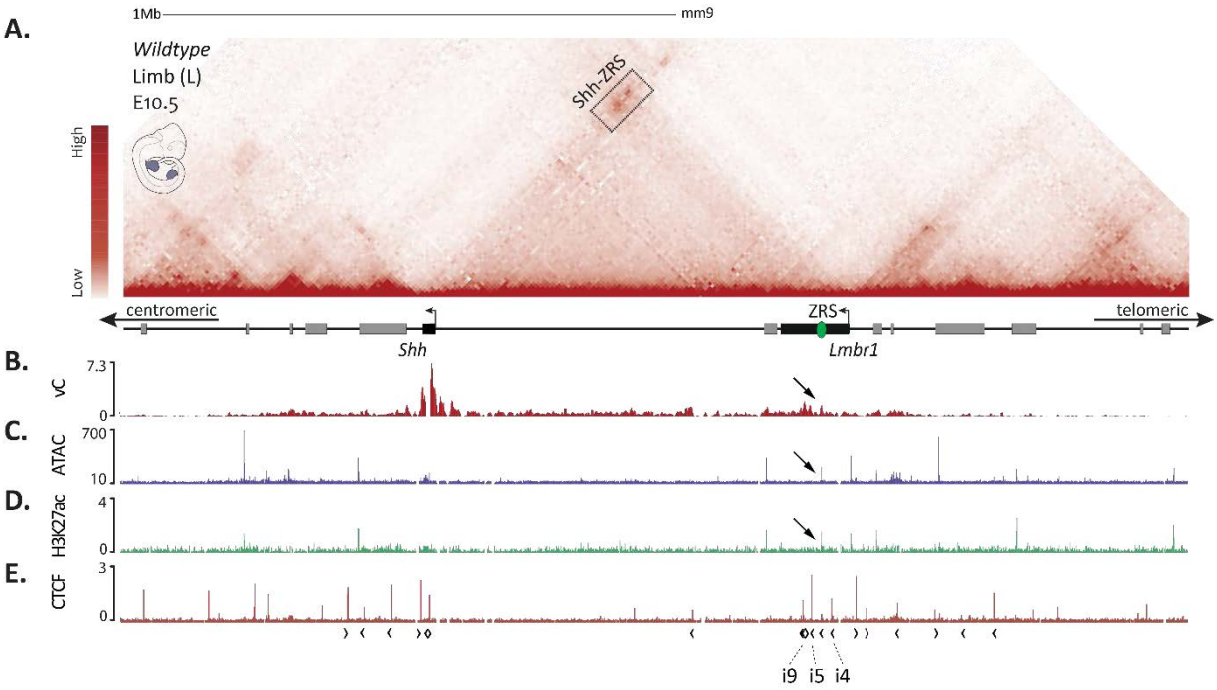


Figure 4. 2. 3D chromatin structure and regulatory landscape of the Shh locus in E10.5 mouse limbs

A. cHi-C map in E10.5 limb buds (forelimbs and hindlimbs). Each pixel of the heatmap reflects the interaction between two DNA fragments, while the different shades of red indicate the interaction frequency. The darker the red the higher the interaction frequency observed. Below, *Lmbr1* and *Shh* genes are indicated as black bars, while neighboring genes are colored grey. The active limb enhancer (*ZRS*) is indicated by a green oval. Note that the black box in the cHi-C map indicates the domain of high interaction between *Shh* and the *ZRS* region. **B.** Virtual capture-C (vC) from the *Shh* promoter highlights frequent chromatin interactions with the *ZRS* (see black arrow). **C.** ATAC-seq (purple) and **D.** H3K27ac (green) tracks in limbs. Black arrows indicate the active limb enhancer, *ZRS*. **E.** CTCF ChIP-seq tracks of the extended *Shh* locus. Black arrows under the ChIP-seq track indicate the orientation of CTCF sites based on FIMO prediction. Note that CTCF motifs i9, i5 and i4 around the ZRS are in convergent orientation to CTCF motifs at the *Shh* region. Figure was adapted from (Paliou et al. 2019).

4.1.2 Tissue-invariant interaction between *Shh* and the *ZRS*

Previous studies have shown that the *ZRS* is found in close proximity to the *Shh* gene in both *Shh*-expressing and non-expressing cells (Amano et al. 2009; Williamson et al. 2016), suggesting that their interaction is rather stable between tissues. To investigate whether and how the interaction observed in limb buds changes across different tissues, we produced cHi-C maps of the *Shh* in mouse ESCs and in E10.5 midbrain (Figure 4. 3). In mESCs, the locus is in a poised transcriptional status, which means that neither the enhancers are active nor the *Shh* gene is expressed. The cHi-C map in mESCs displays a similar overall TAD configuration compared to limb tissue (Figure 4. 3A, C). Moreover, the TAD is delimited by the same genomic boundaries (*Shh* and *Lmbr1*) and similar chromatin looping events are observed at the top corner of the TAD. In E10.5 mouse embryos, *Shh* is expressed in the ventral midbrain and the *zli* region of the brain (Figure 4. 1), and its expression

is under the control of previously described brain *Shh* enhancers, e.g. *SBE1* and *SBE5* (Epstein et al. 1999; Yao et al. 2016). CHi-C heatmap from midbrain tissue is largely comparable with the cHi-C map from limbs and ESCs (Figure 4. 3 B, C). To precisely assess *Shh-ZRS* interactions, virtual capture-C tracks were derived from the corresponding cHi-C heatmaps. Using the *Shh* promoter as viewpoint, a high degree of interaction was observed in all three tissues/cells examined with the *ZRS* region (Figure 4. 3 D).

Furthermore, using ATAC-seq and H3K27ac ChIP-seq, the chromatin properties of the locus were analyzed in the three tissues (Figure 4. 3 E). In mESCs, the chromatin is mostly inaccessible and depleted from active enhancer marks confirming that the locus is inactive. In midbrain, instead, both ATAC-seq and H3K27ac peaks are detected under the *SBE1* MB enhancer in the intron 2 of *Shh* gene and *SBE5* zli enhancer in the intron 15 of *Lmbr1* gene. In contrast to these tissue-specific chromatin states within the *Shh* TAD, CTCF ChIP-seq from mESCs and midbrain revealed a tissue invariant binding profile when compared to limb tissue (Figure 4. 3 F).

Taken together, these results show that the 3D topology at the *Shh* locus is already preset in ESCs. In accordance with previous studies (Amano et al. 2009; Williamson et al. 2016), the interaction between *Shh* and the *ZRS* 3D is largely tissue-invariant, despite the changes in the enhancer repertoire that is being used or the transcriptional status of the locus. However, the factors supporting this preformed configuration are not yet defined.

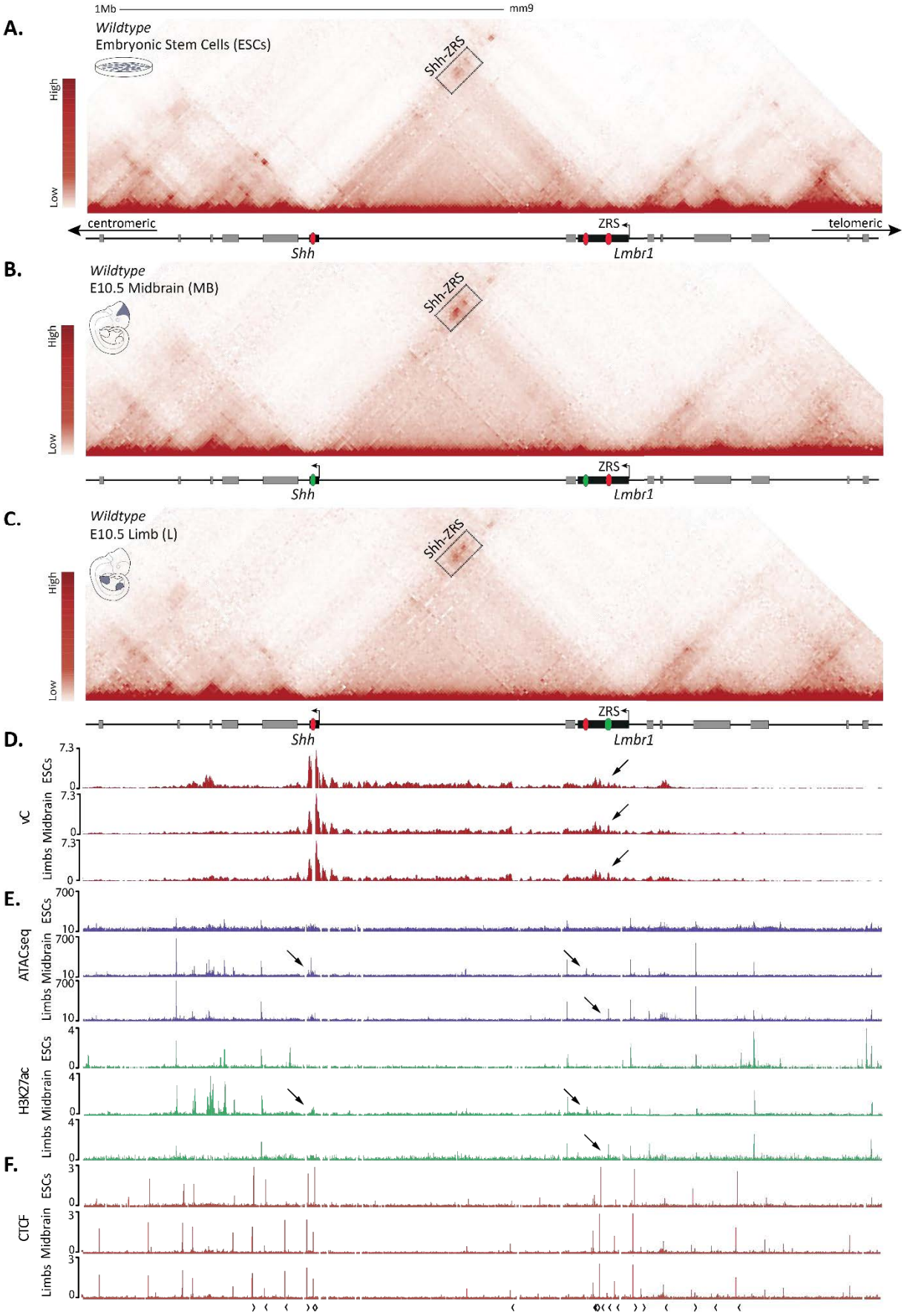

A.
1Mb
mm9
Wildtype
Embryonic Stem Cells (ESCs)
Shh-ZRS
High
Low
centromeric
telomeric
ZRS
Shh
Lmbr1
B.
Wildtype
E10.5 Midbrain (MB)
C.
Wildtype
E10.5 Limb (L)
D.
vC
ESCs
Midbrain
Limbs
E.
ATACseq
H3K27ac
F.
CTCF

Figure 4. 3. Shh and the ZRS enhancer are engaged in a tissue-invariant chromatin interaction.

A. cHi-C map of the extended *Shh* locus in mouse Embryonic Stem Cells (mESCs). Midbrain and limb inactive enhancers are indicated by red ovals on the lower gene track. **B.** cHi-C map in E10.5 midbrain (MB). The inactive *ZRS* enhancer is indicated by a red oval, the active MB and zli enhancers (*SBE1, SBE5*) in the transcribed *Shh* and *Lmbr1* gene are indicated by green ovals. **C.** cHi-C map in E10.5 limb buds (forelimbs and hindlimbs). The active limb enhancer (*ZRS*) is indicated by a green oval, the inactive enhancers are indicated by red ovals. The black boxes in A, B and C indicate the domain of high interaction between *Shh* and the *ZRS* region. Note that the contact does not change in the three cHi-C maps. *Lmbr1* and *Shh* genes in A-C are indicated as black bars, while neighboring genes are colored grey. **D.** Virtual capture-C (vC) from the *Shh* promoter shows similar interactions with the *ZRS* (black arrows) between the three tissues. **E.** ATAC-seq (purple) and H3K27ac (green) tracks in ESCs (upper), midbrain (middle) and hindlimbs (lower tracks). Black arrows indicate the active limb and midbrain enhancers. **F.** CTCF ChIP-seq tracks of the extended *Shh* locus. Note the absence of changes between the three tissues. Black arrows under the ChIP-seq track indicate the orientation of CTCF sites. Figure was adapted from (Paliou et al. 2019).

4.2 Role of constitutive transcription in the preformed architecture at the *Shh* locus

Active transcription at TAD boundaries has been reported by several studies (Dixon et al. 2012; Rao et al. 2014; Bonev et al. 2017), though its exact function in relation with TAD formation is yet to be elucidated. Additionally, Pol II has been shown to be involved in preformed topologies by maintaining loci at a poised status (Ghavi-Helm et al. 2014). At the *Shh* locus, the telomeric boundary, containing the *ZRS*, extends across the *Lmbr1* gene which is constitutively transcribed. We thus postulated that the process of *Lmbr1* transcription could participate in shaping the 3D structure of the locus.

4.2.1 Disruption of constitutive transcription by CRISPR-Cas9

To investigate whether the position of the *ZRS* within a constitutively transcribed gene plays a role in the establishment of the preformed locus topology we abolished *Lmbr1* transcription using CRISPR-Cas9 in ESCs. To do so, we designed two gRNAs (Table 9. 1) targeting a region flanking the *Lmbr1* TSS and promoter based on the ATAC-seq, H3K4me3 and H3K4me1 ChIP-seq signals, known to mark active promoters (Figure 4. 4 A).

An ESC clone with a homozygous *Lmbr1* promoter deletion ($Lmbr1^{\Delta prom/\ \Delta prom}$) of ~2.3 kb was identified and used for tetraploid complementation to generate transgenic animals (Kraft et al. 2015). The genotype was assessed by genotyping PCR, copy number detection by qRT-PCR and the breakpoint of the deletion was validated by Sanger sequencing in targeted ESCs and mutant mouse tissue. To validate the loss of transcription across *Lmbr1*, RNA-seq and ChIP-seq for the transcription elongation histone mark H3K36me3 were performed on somite-staged E10.5 wildtype and $Lmbr1^{\Delta prom/\ \Delta prom}$ limb buds (Figure 4. 4 B). RNA-seq revealed the absence of *Lmbr1* RNA transcript in $Lmbr1^{\Delta prom/\ \Delta prom}$ mutants. H3K36me3 ChIP-seq confirmed the complete loss of transcriptional

elongation over the gene body compared to wildtype limbs. In addition, ATAC-seq in *Lmbr1*$^{\Delta prom/\ \Delta prom}$ mutants was performed to analyze potential changes induced by the absence of *Lmbr1* transcription (Figure 4. 4 B). The comparison to wildtype tissue revealed the complete abrogation of any signal at the *Lmbr1* promoter in *Lmbr1*$^{\Delta prom/\ \Delta prom}$ mutants. However, no additional changes in chromatin accessibility were observed at the locus.

This genetic approach, though removing transcriptional activity, leaves intact the genomic sequence (except 2.3kb) that corresponds to the telomeric boundary of the *Shh* TAD. It is thus interesting to evaluate the potential influence on the 3D structure, but also its eventual effect on *Shh* expression.

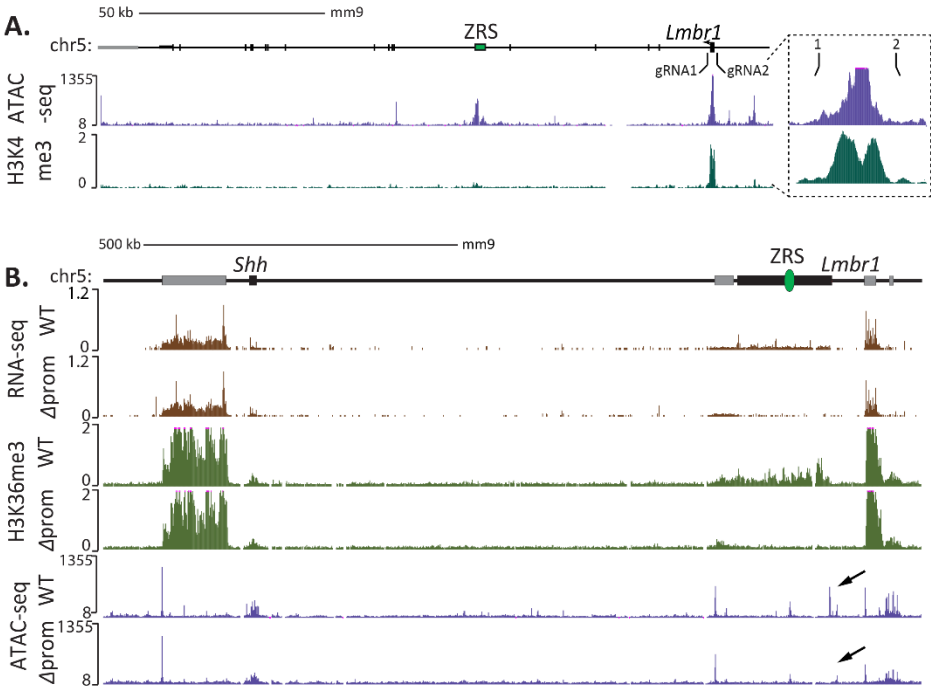


Figure 4. 4. Disruption of Lmbr1 transcription by CRISPR-Cas9

A. ATAC-seq (blue) and H3K4me3 ChIP-seq (green) signal at the *Lmbr1* gene in wildtype E10.5 limb buds. Zoom-in shows *Lmbr1* promoter and position of two flanking sgRNAs for the targeted deletion. **B.** RNA-seq (brown), H3K36me3 ChIP-seq (green) and ATAC-seq (purple) of wildtype and *Lmbr1*$^{\Delta prom/\Delta prom}$ E10.5 limb buds. H3K36me3 mark is removed from the *Lmbr1* gene body, where the RNA transcript is absent. Note the complete loss of ATAC-seq signal at the region of the promoter due to its deletion (black arrows). B was adapted from (Paliou et al. 2019).

4.2.2 Disruption of constitutive transcription leads to differential chromatin interactions

In order to assess if *Lmbr1* transcription has an active role in the formation of the *Shh* TAD and its boundaries, cHi-C was performed in *Lmbr1*$^{\Delta prom/\ \Delta prom}$ mutant limb buds and compared to wildtype. We observed that the overall TAD is to a great extent identical between the two maps (Figure 4. 5

A). However, in *Lmbr1*$^{\Delta prom/\Delta prom}$ limbs, a different interaction frequency was observed between *Shh* and the centromeric part of *Lmbr1*. To better visualize this difference, a subtraction map of *Lmbr1*$^{\Delta prom/\Delta prom}$ *versus* wildtype cHi-C map was produced (Figure 4.5 A, third panel). In this particular map, we could observe higher interactions between *Shh* and sequences centromeric to the *ZRS*.

To validate this slight shift of interactions, 4C-seq was used as an orthogonal approach (Figure 4. 5 B). Here, we used the *Shh* promoter as viewpoint to detect all its interaction partners at the locus. The 4C signal recapitulated robustly the cHi-C results, showing that *Shh* is found in close proximity more frequently with the *ZRS* itself, as well as fragments that directly surround it. Moreover, subtracting wildtype from mutant 4C signals further showed, in *Lmbr1*$^{\Delta prom\ \Delta prom}$, a gain of interactions between *Shh* and the centromeric part of the *Lmbr1* simultaneously with a loss of interactions with the telomeric part of the gene, where *ZRS* is located (Figure 4. 5B, third panel). We then compared the interaction changes observed in *Lmbr1*$^{\Delta prom/\ \Delta prom}$ mutants with wildtype CTCF ChIP-seq tracks (Figure 4. 5 C). We could observe that the centromeric *Lmbr1* gain-of-interaction corresponds to the CTCF binding events in intron 9 (i9) and 5 (i5), while the loss-of-interaction occurred on the telomeric CTCF i4 binding site.

These data demonstrate that abrogation of active transcription over the *ZRS*-containing TAD boundary does not have an effect on global TAD formation. Nevertheless, loss of constitutive transcription modulates *Shh* interactions within the *Lmbr1* gene, possibly leading to a *Shh-ZRS* mis-communication.

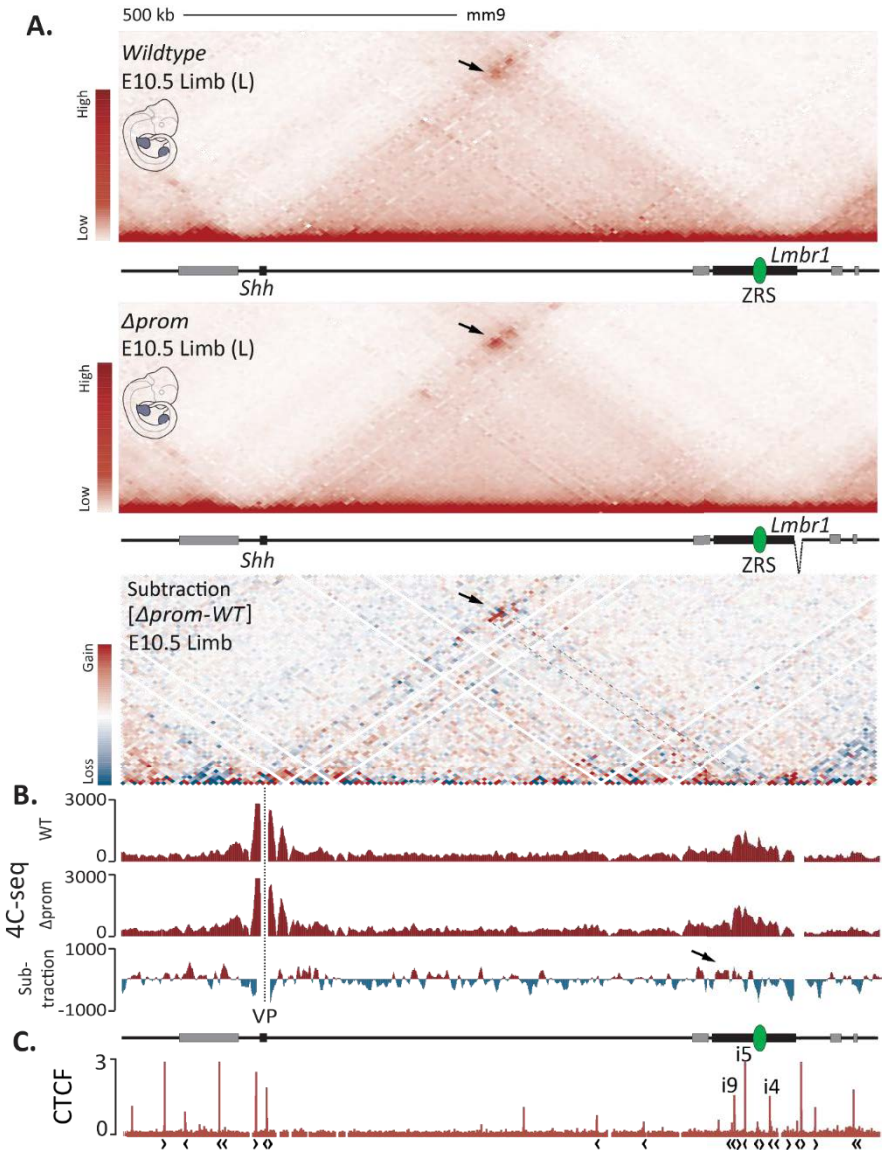


Figure 4. 5. Redistribution of chromatin contacts in Lmbr1$^{\Delta prom/\Delta prom}$ mutants

A. cHi-C maps of wildtype, *Lmbr1*$^{\Delta prom/\ \Delta prom}$ and subtraction map (lowest panel) of *Lmbr1*$^{\Delta prom/\Delta prom}$ *versus* wildtype E10.5 limb buds. Black arrows indicate the differential interaction of the *Shh-ZRS* region. Note that the overall structure does not change between the wildtype and mutant limb buds. **B.** 4C-seq tracks from the *Shh* promoter as viewpoint (VP) in wildtype and *Lmbr1*$^{\Delta prom/\ \Delta prom}$ E10.5 limb buds. The lowest track shows the subtraction of both signals (blue and red indicate loss or gain of contact, respectively, in mutant compared to the wildtype). The black arrow indicates the increase of interaction between *Shh* and the centromeric part of the *Lmbr1* gene in the *Lmbr1*$^{\Delta prom/\ \Delta prom}$ mutants. **C.** CTCF ChIP-seq track of the extended *Shh* locus. The CTCF sites i4, i5 and i9 are indicated. Black arrows under the ChIP-seq track indicate the orientation of CTCF sites. Figure was adapted from (Paliou et al. 2019).

4.2.3 Disruption of constitutive transcription results in decreased *Shh* transcription

The process of transcriptional elongation by Pol II unwinds double-stranded DNA and displaces nucleosomes. Therefore, we reasoned that the function of the *ZRS*, which locates within the constitutively transcribed *Lmbr1* gene body, might be affected by such a process. Specifically, we postulated that transcriptional elongation might sustain a more open chromatin state and easier

binding of transcription factors to the chromatin. To check whether the accessibility of the *ZRS* or, more generally, its activity is affected by the disruption of the *Lmbr1* transcription, ATAC-seq and H3K27ac ChIP-seq were used (Figure 4. 6 A). The ATAC-seq signal at the *ZRS* is largely unchanged compared to wildtype, suggesting that it is still accessible and has not been affected by loss of transcription at the locus. Similarly, the *ZRS* is marked by H3K27ac in *Lmbr1*$^{\Delta prom/\ \Delta prom}$ limb buds suggesting that its activity is unaltered.

To examine the effect of abolishing *Lmbr1* transcription, we performed WISH on *Lmbr1*$^{\Delta prom/\ \Delta prom}$ embryos (Figure 4. 6 B). The comparison of mutant to wildtype stained limb buds did not reveal changes in the distal-posterior localization of *Shh* mRNA. As WISH is only a qualitative measure of expression, qRT-PCR was performed in limb buds to measure eventual changes in the abundance of *Shh* transcription. *Lmbr1*$^{\Delta prom/\ \Delta prom}$ mutants exhibited a 20% decrease in *Shh* transcription. This modest reduction did not lead to any phenotypic alterations in E18.5 *Lmbr1*$^{\Delta prom/\ \Delta prom}$ mouse embryos, which were inspected for phenotypical changes in the skeletal system (data not shown). Taken together, *Lmbr1*$^{\Delta prom/\ \Delta prom}$ limb buds display a moderate loss of *Shh* transcription, which does not result from the enhancer's incapability to function. However, based on the cHi-C data, the observed shift of interactions towards the i9 and i5 CTCF sites suggests a decreased interaction with the *ZRS* and the i4 CTCF site. It is thus plausible that, in *Lmbr1*$^{\Delta prom/\ \Delta prom}$ embryos, differential use of CTCF sites within the telomeric TAD boundary leads to an increased isolation between the *ZRS* and *Shh*.

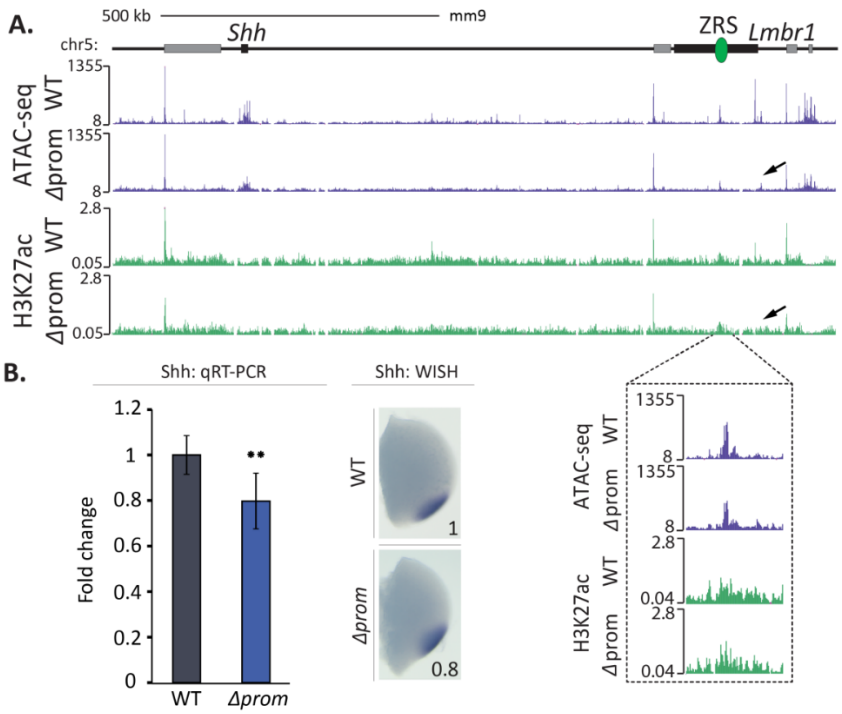


Figure 4. 6 Disruption of Lmbr1 transcription leads to reduced Shh transcription

A. ATAC-seq (purple) and H3K27ac ChIP-seq (green) of wildtype and *Lmbr1*$^{\Delta prom/\Delta prom}$ E10.5 limb buds. In the zoom-in of the H3K27ac and ATAC-seq tracks, note that the *ZRS* enhancer remains open and active (dashed box). **B.** qRT-PCR of *Shh* in wildtype (n=5) and *Lmbr1*$^{\Delta prom/\ \Delta prom}$ E10.5 limb buds (n=3). The p-value was calculated using a one-sided student t-test, where p*<0.05, p**<0.01 and p***<0.001. Error bars represent standard deviation (SD). On the right panel, WISH in wildtype and *Lmbr1*$^{\Delta prom/\ \Delta prom}$ E10.5 limb buds. The numbers correspond to the mean expression level determined by qRT-PCR. Figure was adapted from (Paliou et al. 2019).

4.3 Role of CTCF binding in the 3D architecture at the *Shh* locus

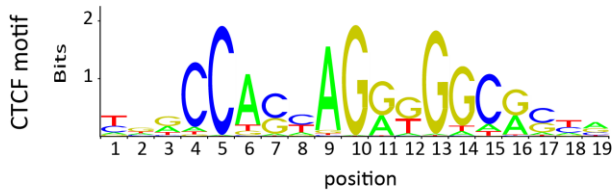
Shh and the *ZRS* are both located between CTCF sites that are in opposite orientation to one another that could define the TAD boundaries. To assess the role of CTCF in the tissue-invariant 3D configuration of the locus, and its role toward *Shh* regulation, CTCF binding sites were deleted and the changes in the structure and regulation were subsequently assessed in embryonic limb buds.

4.3.1 Disruption of CTCF binding motifs by CRISPR-Cas9

According to the cHi-C data (Figure 4. 2, Figure 4. 3), the looping events observed at the *Shh* locus occur between the CTCF sites found on each side of the *Shh* gene and the CTCF sites on each side of the *ZRS*. Specifically around the *ZRS*, we observed two close (~20kb) CTCF binding events in intron 4 and 5 (i4 and i5) of *Lmbr1* and three a bit further (~40kb) in intron 9 (i9) in E10.5 limb buds. To test whether CTCF binding is important for *Shh-ZRS* communication, we disrupted all CTCF binding sites within *Lmbr1* using CRISPR-Cas9.

In detail, sgRNAs (Table 9. 1) were designed to target the JASPAR CTCF motifs detected at the i4 and at the i5 binding sites (Figure 4. 7 A). To abrogate the CTCF binding at the i9 site, which comprises three CTCF binding events, a pair of gRNAs was used to create an approximately 5kb deletion (Figure 4. 7 B). Initially, CTCF i4 and i5 were targeted separately. In order to combine motif deletions without altering the intermediate genomic sequence, ESCs with a homozygous deletion for either of the binding sites were retargeted to disrupt an additional CTCF motif. For instance, homozygous *ΔCTCF i5* ESCs were used for targeting the i4 CTCF site, resulting in *ΔCTCF i4:i5* (Figure 4. 7 B). ESC clones were characterized by genotyping PCR, qRT-PCR and Sanger sequencing. To ensure that the CTCF motifs were deleted in both alleles, several PCR products including the targeted motif were cloned into a pTA-GFP vector (Section 9.1) and sequenced to detect individual alleles. Clones with the smallest possible deletion spanning the core motif in both alleles were selected for further experiments. In addition, the absence of sequence alteration in the *ZRS* enhancer itself was verified by sequencing the selected clones.

A.



B.

Genotype	Alleles	Genomic sequence at i4 CTCF (chr5:29662545-29662614)
Wildtype		AATTACTGGCTCACCATAGAGTTATCTTCTA**CTGCCACCTGCTGGTATCA**CAAAATAGTTCAAAATACAA
ΔCTCF i4	14bp	AATTACTGGCTCACCATAGAGTTATCTTCTA**CTGCCA**----------AAAATAGTTCAAAATACAA
	24bp	AATTACTGGCTCACCATAGAGTTA----------**CA**CAAAATAGTTCAAAATACAA
ΔCTCF i5		Wildtype
ΔCTCF i4:i5, ΔCTCF i4:i5:ZRS, ΔCTCF i4:i5:ZRS:i9	62bp	AATTA----------CAA
	27bp	AATTACTGGCTCACCATAGAGTTA----------AAAATAGTTCAAAATACAA

Genotype	Alleles	Genomic sequence at i5 CTCF (chr5:29619921-29619970)
Wildtype		CCTAGCAGCCTCTGG**AGTCCTCTAGTGGCCAACT**GGAGAACTGCTTGCGG
ΔCTCF i4		Wildtype
ΔCTCF i5, ΔCTCF i4:i5,	11bp	CCTAGCAGCCTCTGG**AGTCCT**----------**CT**GGAGAACTGCTTGCGG
ΔCTCF i4:i5:ZRS, ΔCTCF i4:i5:ZRS:i9	29bp	CCTAGCAGCCT----------CTGCTTGCGG

Genotype	Alleles	Genomic sequence at ZRS CTCF (chr5:29641336-29641386)
Wildtype		TCTGGAAAGAAACCAG**TGCTCCCTAGTGGGGGAGA**GCAGAGAGTTCTGATT
ΔCTCF i4		Wildtype
ΔCTCF i5		Wildtype
ΔCTCF i4:i5		Wildtype
ΔCTCF i4:i5:ZRS, ΔCTCF i4:i5:ZRS:i9	21bp	TCTGGAAAGAA----------**AGA**GCAGAGAGTTCTGATT
	18bp	TCTGGAAAGAAACCAG**TG**----------CAGAGAGTTCTGATT

Genotype	Alleles	Genomic sequence at i9 CTCFs (chr5:29597364-29602779)
Wildtype		**ACAGTGACATCAAGGGACTA**(chr5:29597566-29597585) ,**AGTGCCACCTGTTGGCTGAA** (chr5:29601478-29601497), **GGCCAGAAGAGGGTGTCAG** (chr5:29602087-29602105)
ΔCTCF i4		Wildtype
ΔCTCF i5		Wildtype
ΔCTCF i4:i5		Wildtype
ΔCTCF i4:i5:ZRS		Wildtype
ΔCTCF i4:i5:ZRS:i9	5.8kb	chr5:29596970-29602797
	5.4kb	chr5:29597358-29602836

Figure 4. 7. Genomic deletions of CTCT motifs by CRISPR-Cas9

A. CTCF profile based on JASPAR core database, used for FIMO motif discovery at the *Shh* locus. **B.** Allelic series of CTCF motif deletions. Each panel shows the wildtype genomic sequence of each CTCF site with genomic coordinates and the core motif (in bold). The acquired deletions in both alleles of each CTCF mutant clone are indicated by the gap in the alignment. Note that CTCF deletions differ between *ΔCTCF i4* and *ΔCTCF i4:i5* due to retargeting of the i4 motif in *ΔCTCF i5* ESCs.

4.3.2 Disruption of CTCF sites increases binding on neighboring sites

The above described alleles bear deletion of the CTCF core motif, but in some cases, the resulted deletions span over either a part of the motif or disrupt some bps upstream or downstream in addition to the core motif. Therefore, to validate that the acquired deletions indeed disrupt the binding of CTCF *in vivo*, ChIP-seq for CTCF was performed on mutant E10.5 limb buds and was compared to wildtype limb buds (Figure 4. 8 A, B).

First, *ΔCTCF i4* embryos exhibited loss of CTCF binding at the i4 site, while no additional changes were observed at the locus (Figure 4. 8 B). Similarly, the *ΔCTCF i5* allele successfully abrogated CTCF binding at the i5 site. Surprisingly, in this latter allele, additional CTCF binding was detected at a neighboring binding site at the centromeric edge of the *ZRS*, named as "ZRS CTCF" as well as at a

site close to the *Lmbr1* promoter, named as "Lprom CTCF". The combined deletion of CTCF i4 and i5 in *ΔCTCF i4:i5* mutants showed loss of CTCF binding on the deleted sites, but also an increase at the same sites as in *ΔCTCF i5* limb buds. This observation suggested that the increase of CTCF binding at both sites could potentially compensate the loss of binding at the i4 and i5 targeted sites. As the ZRS CTCF site is found closer to the *ZRS* and more specifically, within the previously suggested long-range regulatory module of the enhancer, it seemed more "potent" to influence the 3D structure (Lettice et al. 2014) (Figure 4. 12, Figure 4. 16). Therefore, we targeted this site in *ΔCTCF i4:i5* ESCs. Interestingly, triple deleted *ΔCTCF i4:i5:ZRS* ChIP-seq showed not only disruption of binding on the deleted sites, but also reduced enrichment on the Lprom CTCF site and the i9 sites found ~40 kb away from the *ZRS*. Finally, to abolish all possible compensatory sites within the *Lmbr1* gene body, the i9 sites were also removed and ChIP-seq results confirmed the complete loss of CTCF binding events (Figure 4. 8B).

As shown in several studies, TAD structures and loops are facilitated by CTCF together with the cohesin complex. We thus next investigated whether CTCF deletions have an effect on cohesin binding at the region. ChIP-seq for the cohesin subunit RAD21 was applied in *ΔCTCF i4:i5* and *ΔCTCF i4:i5:ZRS* limb buds (Figure 4. 8 C). In *ΔCTCF i4:i5* mutants, RAD21 binding follows exactly the same pattern as CTCF binding. Absence of RAD21 is observed at the deleted CTCF sites, while neighboring sites appear to be ectopically bound. However, in *ΔCTCF i4:i5:ZRS* mutants, disruption of three CTCF binding sites was sufficient to abolish RAD21 not only from the deleted sites, but also from the i9 and Lprom CTCF ectopic sites despite that their sequence was intact and still bound by CTCF.

Altogether, these results suggest a compensatory mechanism between CTCF sites that ensures the integrity of the 3D structure. Moreover, the complete loss of cohesin binding upon deletion of the main CTCF sites at the *Lmbr1* locus suggests a cooperative function between these two proteins at the level of recruitment and binding.

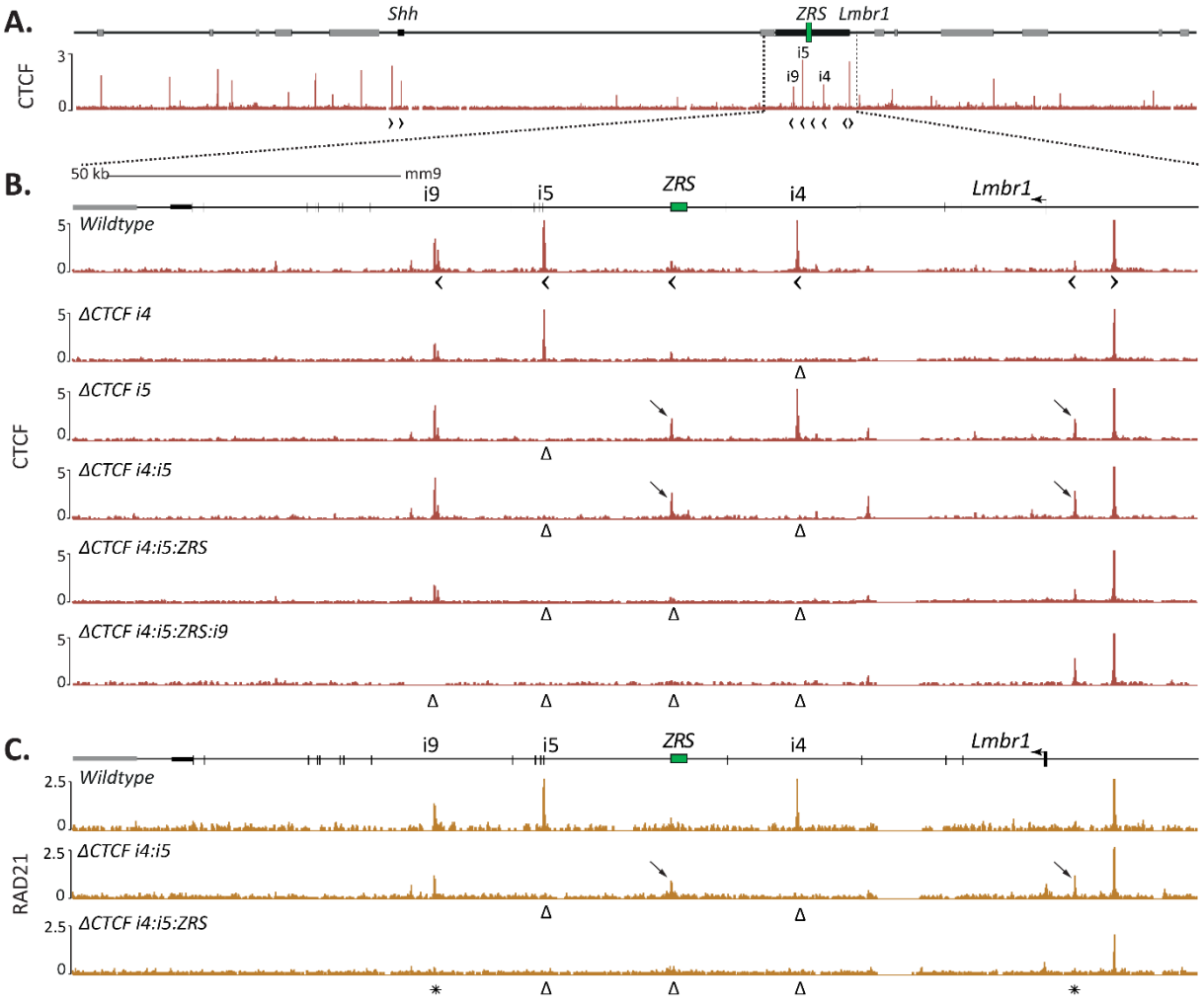


Figure 4. 8. CTCF deletions lead to increased binding on neighboring sites

A. CTCF ChIP-seq enrichment in wildtype E10.5 limb buds at the *Shh* locus. Note the i4, i5 and i9 CTCF binding sites around the *ZRS*. **B.** Zoom-in at the *Lmbr1* gene and CTCF ChIP-seq enrichment in wildtype, *ΔCTCF i4*, *ΔCTCF i5*, *ΔCTCF i4:i5*, *ΔCTCF i4:i5:ZRS* and *ΔCTCF i4:i5:ZRS:i9* homozygote mutants. Δ signs indicate the CTCF motif deletions leading to loss of CTCF binding and the arrows indicate the increased binding on ectopic neighboring sites. **C.** RAD21 ChIP-seq in wildtype, *ΔCTCF i4:i5* and *ΔCTCF i4:i5:ZRS* homozygote mutants. Δ signs indicate the CTCF motif deletions. Black arrows show the increased RAD21 binding on the same, as in CTCF ChIP, ectopic sites. Black asterisks indicate the cooperative RAD21 loss at the i9 and Lprom in *ΔCTCF i4:i5:ZRS* mutants. Figure was adapted from (Paliou et al. 2019).

4.3.3 Loss of CTCF binding alters 3D chromatin structure at the *Shh* locus

Deletion of certain CTCF motifs within *Lmbr1* triggered ectopic CTCF/cohesin binding on alternative neighboring sites suggesting a compensation mechanism that contributes to maintain the 3D architecture of the locus in the nucleus. To assess whether the changes of CTCF binding affect *Shh-ZRS* chromatin interactions and the architecture of the locus, 4C-seq and cHi-C were deployed.

4.3.3.1 *Using 4C-seq to visualize Shh interactions in CTCF mutant limb buds*

In agreement with cHi-C experiments, 4C-seq on wildtype limb buds using the *Shh* promoter as viewpoint showed high interaction frequencies with the *Lmbr1* intragenic region and decreased signal telomeric of the *Lmbr1* promoter (Figure 4. 9 A). In particular, the strongly interacting fragments were located between *Lmbr1* intron 4 and intron 9 and included the CTCF binding sites i4, i5, i9 and the *ZRS* enhancer (Figure 4. 9 A).

To test whether the interaction profile at the locus changes upon disruption of the CTCF motifs, 4C-seq on CTCF mutant limb buds was performed (Figure 4. 9 B). The 4C-seq profile *in ΔCTCF i4* mutants appeared very similar to wildtype 4C-seq, but with a reduced interaction peak at the i4 site. To better visualize the differences we calculated a log2 ratio between wildtype and *ΔCTCF i4* normalized reads. This ratio track clearly showed a loss of interaction at the i4 site, while the rest of the profile remained largely unchanged. By contrast, deleting the i5 CTCF site led to a shift of interactions to the more telomeric side of *Lmbr1*. In particular, loss of signal was observed at the i5 site, while interactions were increased with the *ZRS* and the i4 site. Finally, in double *ΔCTCF i4:i5* mutant limbs, the interactions between *Shh* and the *Lmbr1* intragenic region were evenly and strongly reduced (Figure 4. 9B).

In summary, 4C-seq in CTCF mutants revealed changes in the long-range contacts between *Shh* and the genomic regions surrounding the *ZRS*. Moreover, the shift of interactions observed by 4C-seq followed the pattern of alternating CTCF sites used for binding in the *ΔCTCF i5* mutants, shown in Figure 4. 8. This observation suggests a direct role of CTCF binding at the telomeric *Shh* TAD boundary in mediating long-range chromatin contacts between *Shh* and *ZRS*.

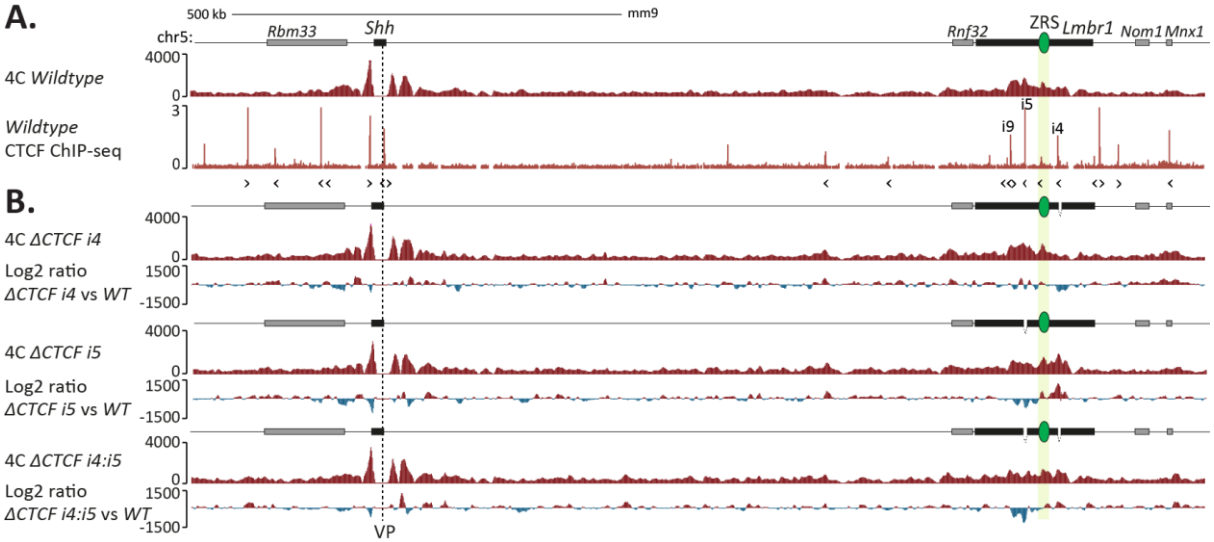


Figure 4. 9. CTCF deletions alter long-range contacts of Shh

A. 4C-seq track from *Shh* promoter as viewpoint (dashed line, VP) in wildtype E10.5 limb buds. The panel below shows CTCF ChIP-seq in wildtype E10.5 limb buds and the arrowheads indicate CTCF orientation. Note that the higher interacting peaks of the profile coincide with *Lmbr1* intragenic CTCF sites (i9, i5, i4) and the *ZRS* enhancer. **B.** 4C-seq tracks from *Shh* promoter as viewpoint (VP) in *ΔCTCF i4, ΔCTCF i5 and ΔCTCF i4:i5* E10.5 limb buds. The panels below each profile show log2 ratio between mutants and wildtype (blue and red indicate loss or gain of contacts in the mutant).

4.3.3.2 Using capture Hi-C to assess the 3D structure in CTCF mutant limb buds

Removal of CTCF motifs might affect not only *Shh*-specific interactions, but many chromatin interactions within the *Shh* TAD. To test that, cHi-C was used as an alternative approach to 4C-seq. As the 4C in *ΔCTCF i4:i5* mutants showed the most changes among the various examined mutant alleles (Figure 4. 9), cHi-C was performed on *ΔCTCF i4:i5* limbs. Comparison of the the *ΔCTCF i4:i5* limb cHi-C map to wildtype map revealed a complete loss of the anchor loops at the top of the *Shh* TAD (Figure 4. 10 A). This effect was particularly visible when subtracting wildtype from *ΔCTCF i4:i5* cHi-C maps (Figure 4. 10 B). Thus, the loss of interactions between *Shh* and the *Lmbr1* intragenic region occurs along with a contact loss between the convergent CTCF sites residing at both TAD boundaries. Additionally to this loss, increased frequency of interactions was observed between the *Shh* TAD and the neighboring *Mnx1*-containing TAD, indicating reduced inter-TAD insulation. Interestingly, despite the loss of major chromatin loops and ectopic interactions between TADs in the *ΔCTCF i4:i5* mutants, the overall structure of the *Shh* TAD and its separation from the neighboring TADs were maintained.

Then, to address whether the increased CTCF binding on the ZRS site detected in the *ΔCTCF i4:i5* mutants (Figure 4. 8) contributes to the 3D structure, we performed cHi-C in the *ΔCTCF i4:i5:ZRS* mutants. Surprisingly, the *ΔCTCF i4:i5:ZRS* cHi-C heatmap was almost identical to the *ΔCTCF i4:i5* interaction map. Overall, loss of interaction was observed between the two TAD boundaries and increased interactions were visible between the two neighboring TADs (Figure 4. 10 D). No additional looping events or changes in the 3D structure were observed.

In order to visualize the difference in the 3D structure of the *ΔCTCF i4:i5* mutants, a conformation prediction approach based on polymer physics was used. Based on the cHi-C data, a 3D model of the locus architecture was produced (Figure 4. 10 E). In the wildtype model, *Shh* and the *ZRS* were found in close proximity and separated from the *Mnx1* gene. In contrast, in *ΔCTCF i4:i5* mutant limb buds, the distance between *Shh* and the *ZRS* was increased and *Mnx1* was found closer to both *ZRS* and *Shh*, in agreement with the observed de-insulation between the *Mnx1* and *Shh* TADs (Figure 4. 10 A, B). The increased *Shh-ZRS* distance was further confirmed by a shift in the distribution of distances across all the polymer-based models derived from wildtype and mutant limb buds (Figure 4. 10 F). Taken together, these data show that the disruption of CTCF binding can destabilize the contacts between the boundaries of the *Shh* TAD, which include *Shh* and the *ZRS*, but is not

sufficient to lead to a complete TAD disruption.

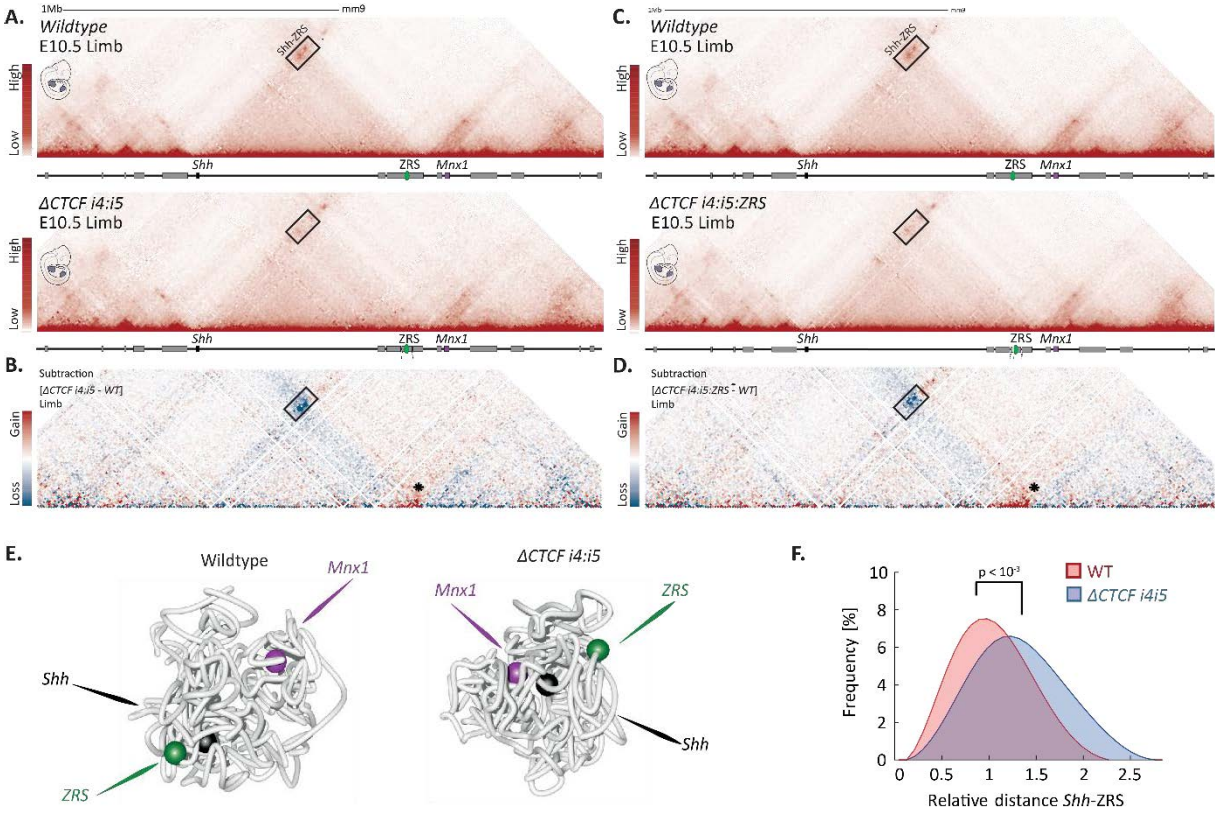


Figure 4. 10. CTCF sites enable the long-range interaction between Shh a

A. cHi-C maps of wildtype (upper) and *ΔCTCF i4:i5* (lower) E10.5 limb buds. **B.** Subtraction map between wildtype and *ΔCTCF i4:i5*, where blue and red indicate loss and gain of contacts, respectively, in the mutants. Black asterisk indicates loss of insulation between *Shh* and *Mnx1* TADs. **C.** cHi-C maps of wildtype (upper) and *ΔCTCF i4:i5:ZRS* (lower) E10.5 limb buds. **D.** Subtraction map between wildtype and *ΔCTCF i4:i5:ZRS*, where blue and red indicate loss and gain of contacts, respectively, in the mutants. Black asterisk indicates the loss of insulation between *Shh* and *Mnx1* TADs. Black box in A to D indicates the domain of high interaction between *Shh* and the *Lmbr1* intragenic region, which contains the *ZRS* enhancer. Note the decreased interaction within the box in *ΔCTCF i4:i5* compared to wildtype tissue. **E.** 3D polymer model of wildtype and *ΔCTCF i4:i5* limb cHi-C data. Note the changes in proximity between *Shh* and *ZRS*, and between *Shh* and *Mnx1*. **F.** Frequency plot of the distance distribution between *Shh* and the *ZRS* in wildtype and *ΔCTCF i4:i5*. Note the increase in relative distance in the mutant limbs. P-value was calculated using the Mann-Whitney test. Figure was adapted from (Paliou et al. 2019).

4.3.4 Expression and phenotypic analysis of CTCF mutants

As shown in Section 4.1.2 , *Shh* and *ZRS* are in close proximity in a tissue-independent manner. Importantly, cHi-C data in combination with 3D polymer-based models showed that the distance between *Shh* and *ZRS* increases significantly upon deletion of both i4 and i5 CTCF sites (Figure 4. 10). To assess how the reduced *Shh-ZRS* distance affects *Shh* regulation, we performed expression analysis in all CTCF mutants in E10.5 limb buds. Moreover, we assessed potential phenotypic outcomes in E18.5 fetuses.

In *ΔCTCF i5* limb buds we measured a 31% loss in *Shh* transcription, while *ΔCTCF i4* limbs displayed a 51% decrease (Figure 4. 11 A). Surprisingly, disrupting both CTCF sites i4 and i5 led to only a marginal additional downregulation of *Shh*. Similarly, and in agreement with the cHi-C data (Figure 4. 10), deleting the ZRS CTCF in the *ΔCTCF i4:i5* background did not further decrease *Shh* transcription levels. Likewise, the complete absence of CTCF sites within the *Lmbr1* region in *ΔCTCF i4:i5:ZRS:i9* mutants did not lead to further decrease in *Shh* transcription. To visualize differences in the pattern of *Shh* mRNA in the limb, WISH experiments were performed in E10.5 wildtype and CTCF mutant embryos (Figure 4. 11 B). Although all wildtype and mutant limbs showed a similar pattern of *Shh* expression, differences were mainly observed as decrease in signal intensity, corresponding to the loss of mRNA measured by RT-qPCR.

To examine whether this loss of transcription affects the phenotype of the mice, E18.5 wildtype and mutant embryos were isolated and skeletal staining was performed (Figure 4. 11 C). Comparison of wildtype with *ΔCTCF i5* limbs revealed no differences in skeletal patterning. By contrast, *ΔCTCF i4* and *ΔCTCF i4:i5* hindlimbs appeared with twisted ankles, resembling a hemimelia phenotype, which molecular cause is not yet understood (Hill et al. 2003). To further investigate the effect of reduced *Shh* expression, *ΔCTCF i4:i5* adult mice were bred with heterozygous for the *ZRS* enhancer, $\Delta ZRS^{+/-}$ mice. The phenotype of *ΔCTCF i4:i5;* $\Delta ZRS^{+/-}$ mutants was examined in detail by comparing the digit length of E18.5 embryos and by performing µ-CT in 3-week old mice (data not shown). This detailed analysis did not reveal skeletal malformations, despite the expected downregulation of *Shh* transcription to ~25%. Nevertheless, it has been reported that phenotypes vary greatly when *Shh* is transcribed in intermediate levels and that 10-55% of *Shh* transcription is sufficient for normal number of digits in the hindlimbs (Krebs et al. 2003; Lettice et al. 2014).

In conclusion, CTCF motif deletions at the telomeric *Shh* TAD boundary lead to a ~50% decrease of *Shh* transcription. However, *Shh* and *ZRS* seem to achieve communication in a CTCF-independent manner, which is sufficient to give rise to normal limbs and digits.

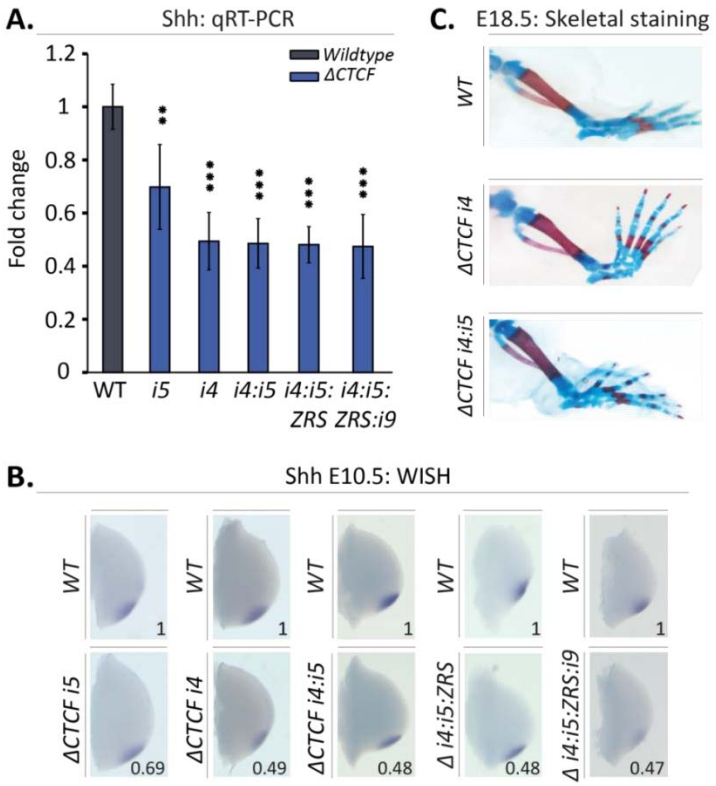


Figure 4. 11. CTCF deletions result in decrease of Shh transcr

A. qRT-PCR of *Shh* in wildtype (n=5), *ΔCTCF i5* (n=3), *ΔCTCF i4* (n=6), *ΔCTCF i4:i5* (n=3), *ΔCTCF i4:i5:ZRS* (n=5) and *ΔCTCF i4:i5:ZRS:i9* (n=4) E10.5 limb buds. Error bars represent standard deviation (SD). The p-value was calculated using a one sided student t-test, where p*<0.05, p**<0.01 and p***<0.001. **B.** *Shh* WISHs in E10.5 wildtype and mutant hindlimb buds. The numbers correspond to the mean expression level determined by qRT-PCR. **C.** E18.5 limb skeletal preparation of wildtype, *ΔCTCF i4* and *ΔCTCF i4:i5* hindlimbs. Note that the mutant hindlimbs are slightly distorted.

4.4 Role of 3D structure in a genetically perturbed *Shh* locus

Although genetic deletions of the CTCF motifs have a strong effect on the chromatin folding at the *Shh* locus and on *Shh* transcription, they were still not sufficient to lead to a loss-of-function phenotype. To determine the role of the 3D structure in more "sensitive" conditions, the locus was further genetically targeted by CRISPR-Cas9.

4.4.1 Partial deletion of the *ZRS* enhancer sequence by CRISPR-Cas9

ZRS enhancer is a highly conserved sequence (Figure 4. 12 A) and has been suggested to be split in two parts with distinct regulatory activities (Lettice et al. 2014). The 5' part of the enhancer defines the spatiotemporal expression of *Shh* in the limb and the 3' part is proposed to play a role in the long-range communication with the *Shh* promoter (Figure 4. 12 B). However, TFBSs necessary for proper *Shh* expression, including ETS and HOX sites, are found across the entire enhancer (Figure 4. 12 B). To test if the 3' part of the enhancer could indeed participate in the CTCF-independent mechanism sustaining the residual *Shh* transcription levels in the CTCF mutants, we engineered its

homozygous deletion using CRISPR-Cas9. Specifically, we designed a pair of sgRNAs (Table 9. 1) to delete the 3'end of the *ZRS* enhancer and obtained a ~400 bp homozygous deletion (*ΔZRSreg*) extending from the ETS2 to the ZRS CTCF site at the centromeric end (Figure 4. 12B).

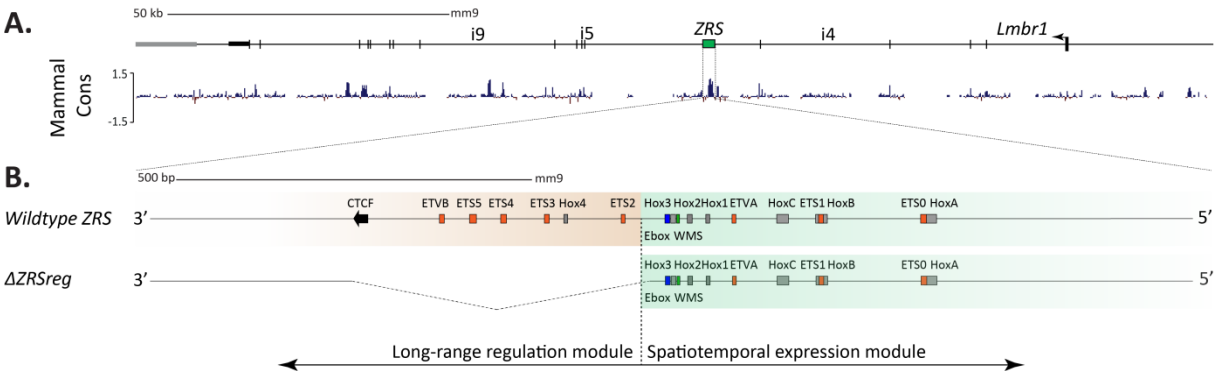


Figure 4. 12. ΔZRSreg mutation removes part of the ZRS enhancer and its binding sites

A. Genomic sequence of the *ZRS* enhancer within intron 5 of the *Lmbr1* gene and conservation across mammals. **B.** Zoom-in in the enhancer sequence and its binding sites. The *ZRS* enhancer accommodates six ETS binding sites, two ETV4/5 sites (orange boxes), seven HOX sites (grey boxes) and one E-box site (blue box) depicted according to the literature terminology (Galli et al. 2010; Lettice et al. 2012; Leal and Cohn 2016; Kvon et al. 2016; Lettice et al. 2017). Note the suggested functionally differential modules (brown and green shades) suggested by Lettice et al. and the CTCF site at the centromeric side of the enhancer (Lettice et al. 2014). *ΔZRSreg* ESCs encompass a ~400 bp deletion of the 3' side of the enhancer.

4.4.2 *ΔZRSreg* mutation leads to weaker *ZRS* activity and oligodactyly

The *ΔZRSreg* deletion disrupts the centromeric half of the *ZRS* enhancer sequence with its respective TF binding sites. If this part of the *ZRS* acts indeed in establishing or maintaining the long-range regulation of *Shh*, its removal should abolish communication with *Shh* and thus, its transcription regulation (Lettice et al. 2014).

To assess potential changes in chromatin organization, cHi-C was performed in *ΔZRSreg* mutant limb buds and compared to wildtype tissue (Figure 4. 13). *ΔZRSreg* mutant limb buds displayed overall the same TAD structure as wildtype limbs (Figure 4. 13 A). In support of this observation, subtraction map between wildtype and *ΔZRSreg* showed no major differences between them (Figure 4. 13 B). However, a modest gain of interactions was observed between *Shh* and the *ZRS* in the *ΔZRSreg* tissue. To visualize in particular the interaction frequency changes with the *Shh* promoter, virtual capture-C profiles were produced (Figure 4. 13 C). Setting *Shh* promoter as viewpoint indeed highlighted some alterations in interaction frequency with sequences around the *ZRS* in the mutant profile. By subtracting the two profiles and aligning all data, we concluded that the gain of interaction frequency occurs with the CTCF sites i4, i5 and i9.

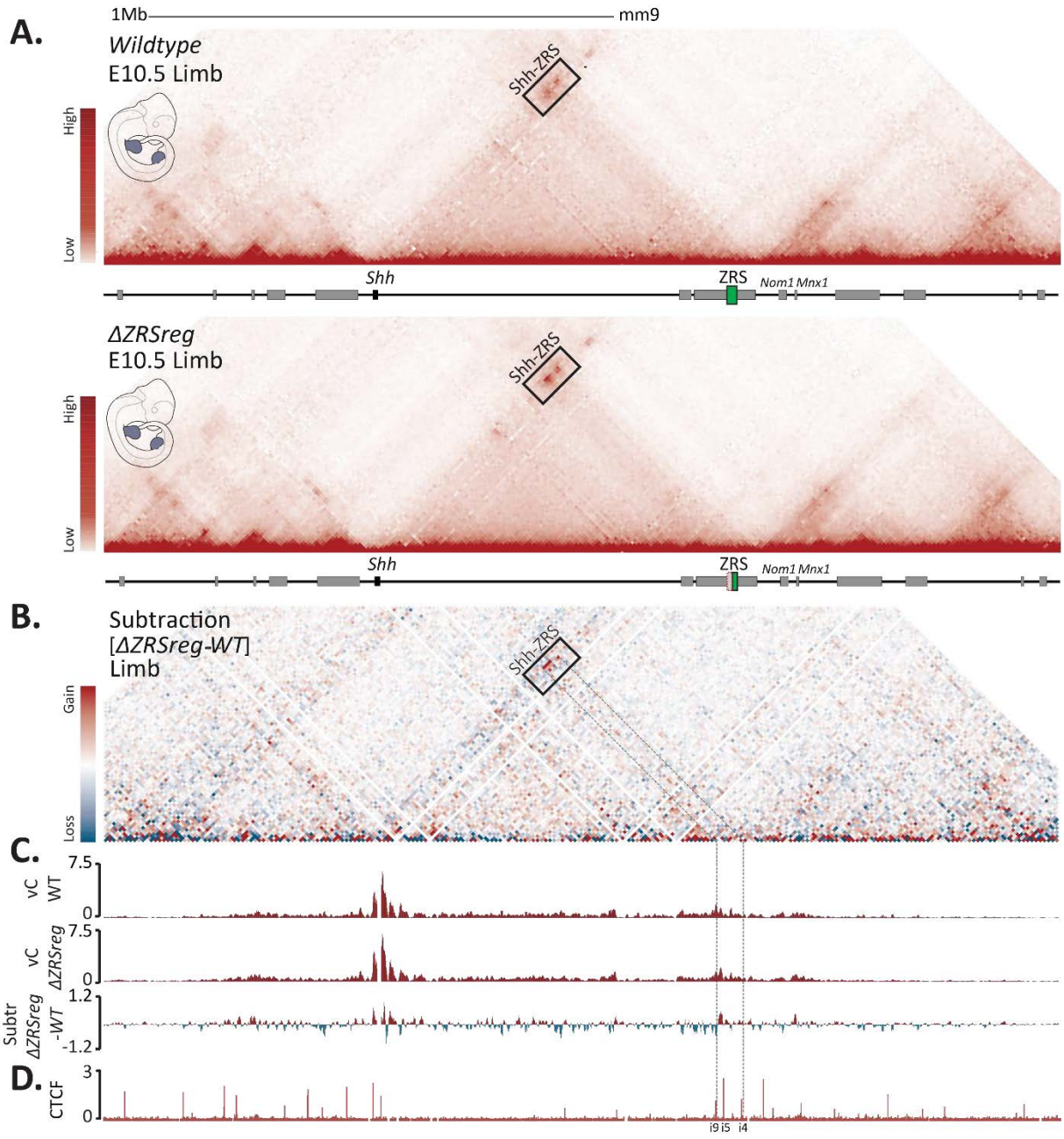


Figure 4. 13. Deletion of 3' module of ZRS enhancer does not alter 3D chromatin structure at the Shh locus

A. cHi-C maps of wildtype (upper) and *ΔZRSreg* (lower) E10.5 limb buds. Black box indicates the domain of high interaction between *Shh* and the centromeric side of *Lmbr1*, which contains the *ZRS* enhancer. **B.** Subtraction map between wildtype and *ΔZRSreg* map, where blue and red indicate loss and gain of contact, respectively, in the mutants. Note that the overall structure does not change between the wildtype and mutant limb buds. A marginal increase at the CTCF sites is indicated in the black box. **C.** Virtual capture-C (vC) profiles using *Shh* promoter as viewpoint derived from cHi-C in E10.5 wildtype and *ΔZRSreg* limb buds. Lower track shows subtraction of the two upper tracks, where positive values indicate gain (red) and loss (blue) of interaction in *ΔZRSreg* limb buds. Note the marginal gain of interaction with the CTCF sites within *Lmbr1* gene. **D.** Wildtype limb CTCF ChIP-seq track. Note the dashed lines aligning the increase of interaction in the cHi-C *ΔZRSreg* heatmap with the i4, i5 and i9 CTCF sites. Figure was adapted from (Paliou et al. 2019).

Despite that the centromeric part of the *ZRS* does not appear to play an essential role in initiating or maintaining the long-range contact with *Shh*, it is expected to affect its regulation. Removal of this long stretch of conserved sequence encompassing important TFBSs might indeed reduce its enhancer activity. Therefore, qRT-PCR was performed in somite-staged E10.5 wildtype and *ΔZRSreg* limb buds. Deletion of the 3' part of the *ZRS* enhancer was sufficient to lead to a significant loss of 75% in *Shh* transcription (Figure 4. 14 A). Reduction of *Shh* mRNA signal was also detected in the WISH experiments performed in E10.5 mouse embryos. This decrease in *Shh* transcription resulted in an oligodactyly phenotype, as detected in skeletal stainings of E18.5 embryos (Figure 4. 14 B). It is noticeable that forelimbs are more affected than hindlimbs, although decrease of *Shh* mRNA can be detected also in the *ΔZRSreg* hindlimb WISH.

Together these data suggest that the 400 bp 3' *ZRS* deletion, previously referred to as the long-range interaction module (Lettice et al. 2014), slightly modifies the interaction profile, but without alteration in the overall 3D structure of the locus. However, this mutation leads to a severe *Shh* loss-of-function with an oligodactyly phenotype, probably due to the deletion of important transcription factor binding sites and subsequent compromised enhancer activity.

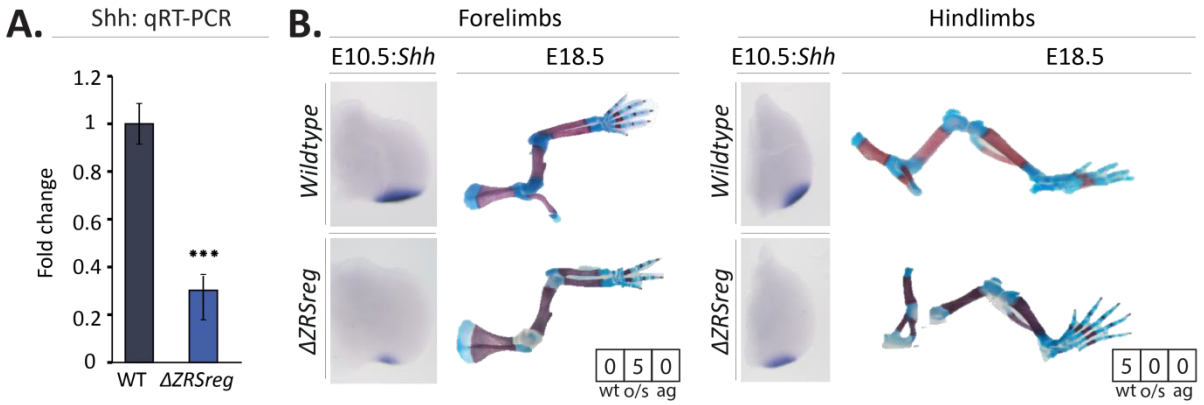


Figure 4. 14. ΔZRSreg mutation induces Shh downregulation and oligodactyly

A. qRT-PCR of *Shh* in wildtype (n=5) and *ΔZRSreg* (n=6) E10.5 limb buds. The p-value was calculated using a one sided student t-test. Error bars represent standard deviation. **B.** *Shh* WISHs in E10.5 limb buds and E18.5 limb skeletal preparation of limbs wildtype and mutants. The boxes represent the number of animals with wildtype (wt), oligodactyly/syndactyly (o/s) or agenesis (ag) phenotype of digits. Note the differences in expression and skeletal defects between wildtype and *ΔZRSreg* mouse limbs. Figure was adapted from (Paliou et al. 2019).

4.4.3 The 3D structure ensures transcription robustness at the *Shh* locus

Deleting CTCF in wildtype conditions led to a significant loss of *Shh* transcription due to the altered 3D chromatin folding (Section 4.3.4). It is intriguing to assess the importance of CTCF in *Shh* regulation by disrupting the 3D structure of the locus in "adverse" genetic conditions. In fact, the

ΔZRSreg allele constitutes a hypomorphic allele with weaker enhancer activity. Therefore, the hypomorphic *ΔZRSreg* mutation was induced in the *ΔCTCF i4:i5* allele by CRISPR-Cas9. The gRNAs used for this targeting were the same as for the generation of the *ΔZRSreg* allele (Table 9. 1). To investigate changes of chromatin structure in the *ΔCTCF i4:i5:ΔZRSreg* mutant limb buds, cHi-C was performed and compared to wildtype (Figure 4. 15). As expected, the chromatin structure of *ΔCTCF i4:i5:ΔZRSreg* limb buds showed, similarly to *ΔCTCF i4:i5* mutants, a strong loss of interaction between *Shh* and the *Lmbr1* boundary as well as a partial de-insulation of the *Shh* TAD (Figure 4. 15 A, B).

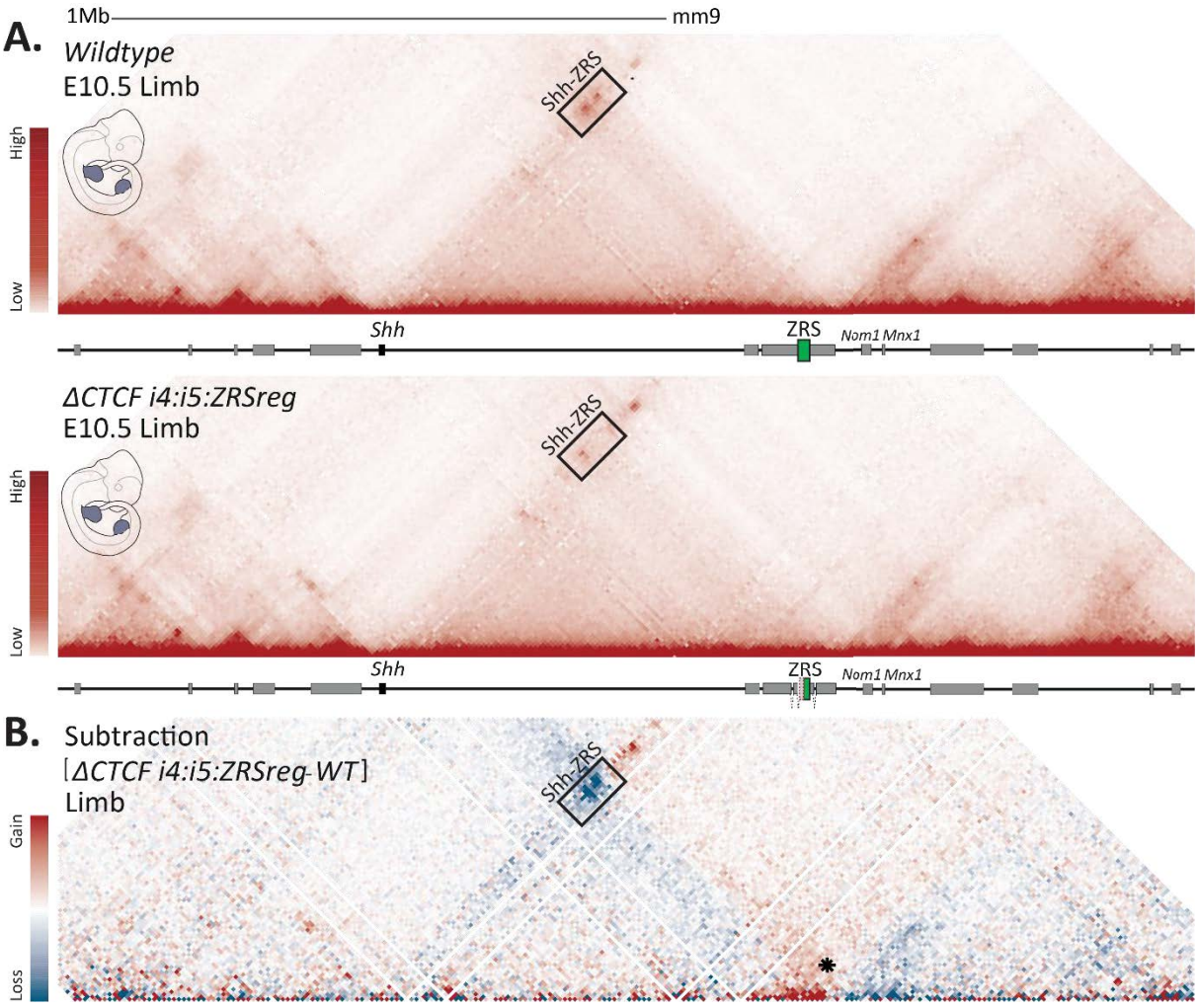


Figure 4. 15. ΔCTCF i4:i5:ΔZRSreg mutation shows loss of interactions and TAD insu

A. cHi-C maps of wildtype (upper) and *ΔCTCF i4:i5: ΔZRSreg* (lower) E10.5 limb buds. Black box indicates the domain of high interaction between *Shh* and the centromeric side of *Lmbr1*, which contains the ZRS enhancer. **B.** Subtraction map between wildtype and *ΔCTCF i4:i5: ΔZRSreg* maps, where blue and red indicate loss and gain of contact, respectively, in the mutants. Black asterisk indicates loss of insulation between *Shh* and *Mnx1* TADs.

In order to better visualize how *Shh*-specific interactions are affected, a virtual capture-C track with the *Shh* promoter as viewpoint was produced from the *ΔCTCF i4:i5:ΔZRSreg* cHi-C map along with *ΔCTCF i4:i5* and *ΔCTCF i4:i5:ZRS* tracks and were compared to wildtype (Figure 4. 16 A). Interestingly, the subtraction profiles showed a great degree of similarity between the three mutants. More specifically, loss of interactions between the two boundaries and gain of interactions with the *Mnx1* TAD were observed in all three experiments (Figure 4. 16 A). Sanger sequencing and ChIP-seq were additionally used to determine differences in sequence and CTCF binding at the *ZRS* between the three mutations (Figure 4. 16 B, C). The *ΔZRSreg* deletion in *ΔCTCF i4:i5:ΔZRSreg* mutants disrupted the sequence of the ZRS CTCF site and therefore, abrogated CTCF binding at this site. This configuration largely resembled the CTCF binding pattern observed in *ΔCTCF i4:i5:ZRS* mutants (Figure 4. 16 B). The only difference between these two alleles is that a 400 bp segment was additionally removed in the *ΔCTCF i4:i5:ΔZRSreg* allowing to determine the contribution of TFBSs to the long-range interaction with *Shh* (Figure 4. 16 C). However, when comparing the two virtual capture-C tracks and their subtraction to the wildtype, we observed a similar decrease in contact frequency with *Shh* (Figure 4. 16 A). Therefore, it seems that loss of the ZRS CTCF and the centromeric part of the *ZRS* in *ΔCTCF i4:i5:ΔZRSreg* does not affect *Shh-ZRS* interactions more strongly than in *ΔCTCF i4:i5:ZRS* mutants.

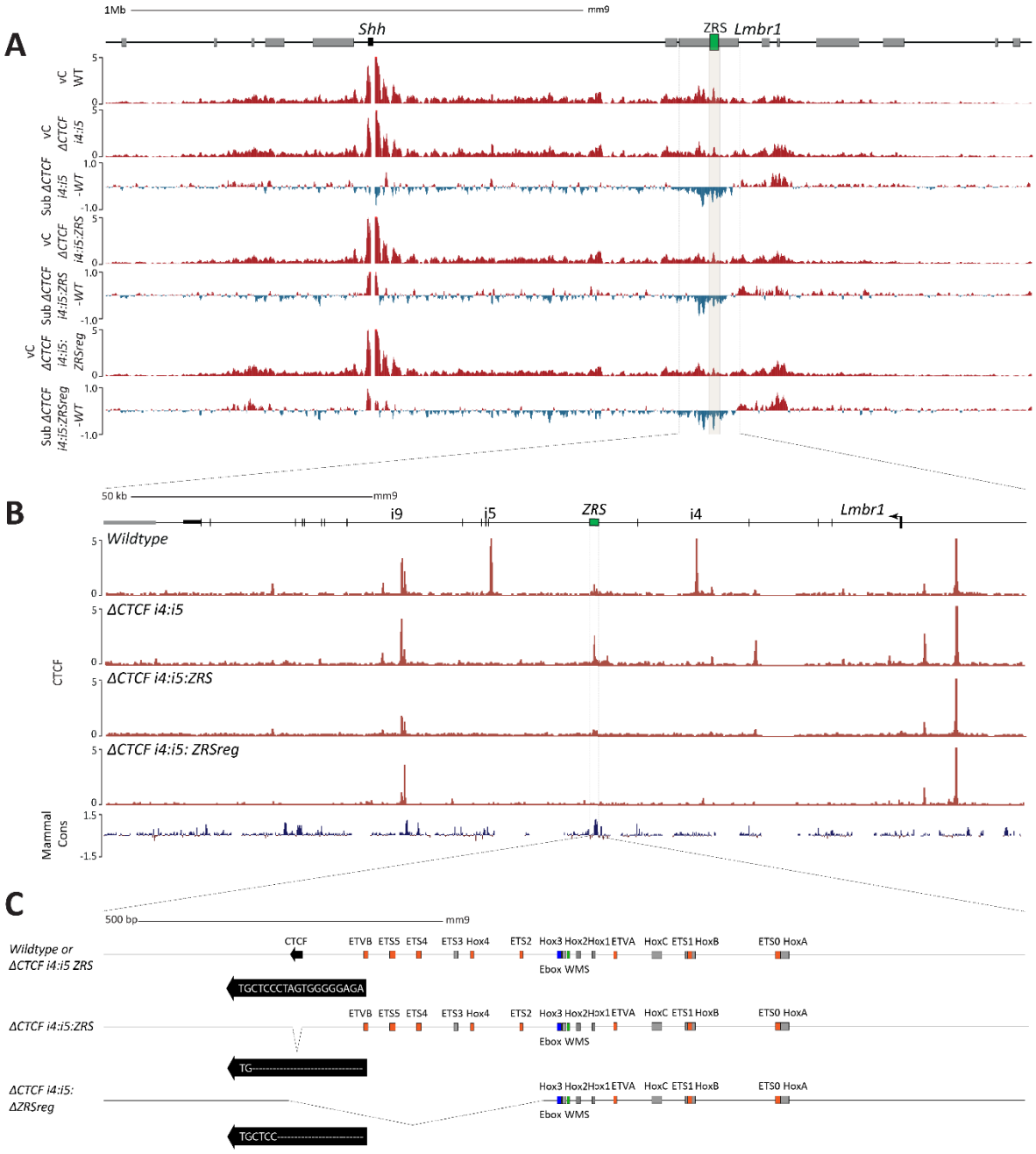


Figure 4. 16. ΔCTCF i4:i5:ZRSreg mutation induces similar changes as ΔCTCF i4:i5 and ΔCTCF i4:i5:ZRS mutations

A. Virtual capture-C (vC) with the *Shh* promoter as viewpoint in wildtype, *ΔCTCF i4:i5*, *ΔCTCF i4:i5:ZRS* and *ΔCTCF i4:i5:ZRSreg* limb buds. Subtraction tracks between mutants and wildtype are presented below each mutant vC track, where red indicates gain and blue indicates loss of interaction in the mutant. The subtraction profiles show decrease of interactions with the CTCF sites within the *Lmbr1* gene and the *ZRS*. Note that the interaction frequency with the *ZRS* enhancer is similar in *ΔCTCF i4:i5:ZRS* and *ΔCTCF i4:i5:ZRSreg* mutants (grey box). **B.** CTCF ChIP-seq tracks of wildtype, *ΔCTCF i4:i5*, *ΔCTCF i4:i5:ZRS* and *ΔCTCF i4:i5:ZRSreg* limb buds. Note that ZRS CTCF binding is absent in both *ΔCTCF i4:i5:ZRS* and *ΔCTCF i4:i5:ZRSreg* mutants. **C.** Zoom-in at the *ZRS* genomic sequence. Known transcription factor binding sites are depicted according to the literature terminology (Galli et al. 2010; Lettice et al. 2012; Leal and Cohn 2016; Kvon et al. 2016; Lettice et al. 2017). ZRS CTCF site sequence is shown in wildtype, *ΔCTCF i4:i5:ZRS* and *ΔCTCF i4:i5:ZRSreg* alleles. The black arrow indicates orientation of the ZRS CTCF site. Figure was adapted from (Paliou et al. 2019).

Next, the impact of the *ΔCTCF i4:i5:ΔZRSreg* mutation on *Shh* transcription was assessed by qRT-PCR and WISH on somite-staged E10.5 mutant and wildtype limb buds (Figure 4. 17). Surprisingly, the mutation resulted in an almost complete absence of *Shh* transcription (> 99%) in limb buds (Figure 4. 17 A). Indeed, *Shh* mRNA signal was not detectable in E10.5 embryos assessed by WISH (Figure 4. 17 B). Similarly to the *ΔZRSreg* mutant embryos, forelimbs were more affected than hindlimbs. Skeletal staining of E18.5 wildtype and *ΔCTCF i4:i5:ΔZRSreg* mutant embryos confirmed the complete loss-of-function phenotype in the forelimbs, characterized by digit agenesis, while hindlimbs showed a fully penetrant oligodactyly phenotype (Figure 4. 17 B).

To better evaluate the degree of remaining *Shh* transcription in *ΔCTCF i4:i5:ΔZRSreg* mutant limbs, an additional allele was engineered that resembled a complete knockout of *Shh* in the limbs. Using the sgRNAs designed centromeric to the CTCF i5 site and telomeric to the *Lmbr1* promoter (Table 9. 1), a homozygous deletion of ~86 kb was induced. In this case, not only the CTCF sites were deleted, but also the complete *ZRS* enhancer (*ΔZRS*). qRT-PCR in *ΔZRS* E10.5 embryos showed complete loss of *Shh* expression and led to fully penetrant digit agenesis phenotype in both E18.5 fore- and hindlimbs (Figure 4. 17).

In summary, in *ΔZRSreg* mutants, where the enhancer's activity is compromised, CTCF-mediated chromatin interactions between *Shh* and the *ZRS* seem to sustain the residual *Shh* expression in the limbs (Figure 4. 14). Accordingly, in *ΔCTCF i4:i5:ΔZRSreg* mice, loss of CTCF is sufficient to abolish *Shh* transcription and lead to loss-of-function phenotype, similar to the *ZRS* KO in limbs. However, the *Shh* transcription level in *ΔCTCF i4:i5:ΔZRSreg* is still higher than in the *ΔZRS* mutants and thereby, suffices for the development of a hindlimb autopod with fewer digits in contrast to the *ΔZRS* zeugopod and autopod agenesis.

Altogether, these data support that the function of the CTCF-mediated preformed structure at the *Shh* locus is to confer robustness in *Shh* regulation and ultimately, to allow for proper digit development.

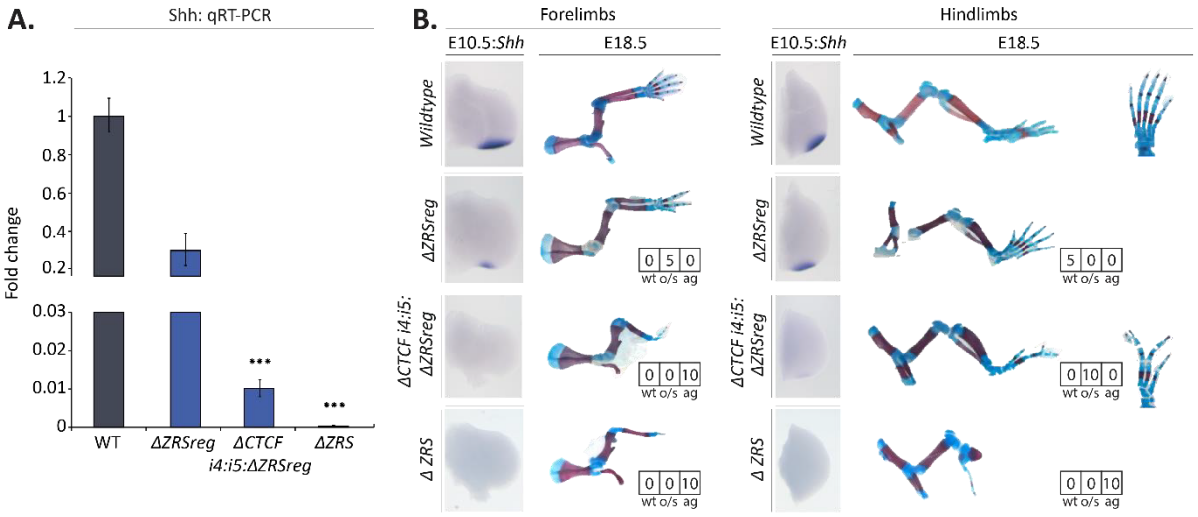


Figure 4. 17. ΔCTCF i4:i5:ΔZRSreg mutation abolishes Shh transcription and leads to digit malformations similar to ΔZRS expression and phenotype

A. qRT-PCR of *Shh* in wildtype (n=5), *ΔCTCF i4:i5:ΔZRSreg* (n=3) and *ΔZRS* (n=3) E10.5 limb buds. The p-value was calculated using a one sided student t-test, where p*<0.05, p**<0.01 and p***<0.001. Error bars represent standard deviation. **B.** *Shh* WISH in E10.5 limb buds and E18.5 limb skeletal preparation of wildtype and mutants. The boxes represent the number of animals with wildtype (wt), oligodactyly/syndactyly (o/s) and agenesis of digits (ag). Note the similarities in expression and skeletal defects between *ΔCTCF i4:i5:ΔZRSreg* and *ΔZRS* mouse limbs. Figure was adapted from (Paliou et al. 2019).

5 DISCUSSION

From bacteria to mammals, the drivers behind chromatin folding are very diverse (Szabo et al. 2019). The development of proximity-ligation assays have allowed to characterize several layers of chromatin organization and modes of enhancer-promoter communication. At some loci, enhancer-promoter proximity in the 3D space of the nucleus is constitutive. Yet, it is not known whether and how it modulates gene regulation. Here, I set out to dissect a tissue-invariant 3D structure and the factors which contribute to its formation using the *Shh* locus as paradigm. By perturbing individual components of the tissue-invariant chromatin architecture and analyzing *Shh* expression *in vivo*, I aimed at assessing the impact of the preformed topology on *Shh* regulation and ultimately, in the development of limbs.

5.1 The *Shh* locus displays a preformed 3D structure

Topologically Associating Domains have been described as architectural units defining regulatory landscapes and delimiting *cis*-interactions between their boundaries (Dixon et al. 2012; Nora et al. 2012; Symmons et al. 2014). At the *Shh* locus, Sleeping Beauty transposons with a *LacZ* reporter gene display staining in the ZPA as long as they are inserted within the *Shh* TAD (Dixon et al. 2012; Anderson et al. 2014). Here, we used Capture Hi-C to enrich for chromatin interactions at the extended *Shh* locus and to produce high resolution heatmaps. Moreover, we performed the experiments in E10.5 mouse limbs, a developmental stage and a tissue in which *Shh* is expressed, in order to assess potential tissue-specific interactions. In agreement with published data in the mouse limb (Williamson et al. 2016), we observed that the *Shh* TAD comprises the *Shh* gene at its centromeric TAD boundary, a long gene desert and the "bystander" genes *Rnf32* and *Lmbr1* at the telomeric TAD boundary. Moreover, chromatin loops were detected at the top corner of the domain, also seen as peaks in the virtual capture-C profiles, representing the most frequently interacting fragments of the locus. These interactions were formed between *Shh* and regions within the constitutively expressed *Lmbr1* gene, which contains the *ZRS* enhancer and several CTCF binding sites.

TADs are suggested to be stable across tissues and species (Dixon et al. 2012), but chromatin loops can be more dynamic (Rao et al. 2014). Here, mESCs-derived cHi-C maps were used as control tissue, where both *Shh* and *ZRS* are inactive. These cHi-C heatmaps contained higher number of valid interacting pairs, which resulted in better resolution of the enriched locus structure compared to mESC published data (Bonev et al. 2017). Then, cHi-C maps from E10.5 midbrain were used as additional control tissue, where *Shh* is expressed but is regulated by different enhancers than the *ZRS*. Comparing the locus in three different transcriptional states revealed that the CTCF-CTCF loops

between the two boundaries that were observed in limbs were very similar and the 3D structure remained globally tissue-invariant. Remarkably, no additional loops were observed in any of the three cell types/tissues. This absence of looping between midbrain enhancers and the *Shh* promoter can be interpreted in two ways; first, the contact might be very transient and cannot be detected by cHi-C and second, enhancer-promoter proximity might not be necessary for gene activation. Supporting the latter case, Benabdallah and colleagues could not observe any chromatin looping between the brain enhancers of the *Shh* locus with the *Shh* promoter in mouse neural progenitor cells (NPCs) (Benabdallah et al. 2017). Oppositely, looping was detected between *ZRS* and *Shh* in the nucleus of mouse limb bud cells (Amano et al. 2009; Williamson et al. 2016). More importantly, our results together with these previous studies show that contact is established between *Shh* and the *ZRS* in *Shh*-expressing, but also in non-expressing cells. In the ZPA though, where the *ZRS* is active, *Shh-ZRS* colocalization is significantly increased (Williamson et al. 2016). This increased colocalization in the limb compared to midbrain or ESCs was not observed in our experiments. One reason could be the limitation of proximity ligation assays to measure frequency of contact between interacting fragments within a range of distance, defined by the crosslinking and ligation efficiency. Super resolution imaging with probes on *Shh* and the *ZRS* could give a better insight on the distribution of distances between them across many nuclei and thus, enable a more precise quantification of distance changes. Moreover, microdissection of the distal posterior limb or ZPA cell sorting would probably increase the precision of the analysis. However, it would be challenging to dissect in a reproducible manner and would require a considerably larger number of sacrificed animals.

Overall, our data demonstrate the characteristics of the *Shh* TAD, which seems to belong to the -initially reported- one third of TADs, which boundaries are accommodated by one housekeeping gene as well as CTCF binding sites (Dixon et al. 2012; Rao et al. 2014; Bonev et al. 2017). Accordingly, these associated characteristics were the obvious candidates for establishing the preformed 3D chromatin architecture of the *Shh* locus.

5.2 CTCF can mediate preformed chromatin topologies

5.2.1 A dynamic interplay between CTCF and cohesin maintains the 3D structure

CTCF occupancy at TAD boundaries is associated with the formation of stable chromatin loops (Rao et al. 2014; Vietri Rudan et al. 2015). At the *Shh* TAD also, the invariable looping events detected by cHi-C occur between CTCF-bound sites. Two CTCF binding sites lie at the centromeric boundary around the *Shh* gene, while three sites are found around the *ZRS* within the telomeric boundary. Centromeric and telomeric CTCF sites, which are also bound by cohesin, display motifs in

convergent orientation suggesting their implication in loop formation *via* the loop extrusion model (Sanborn et al. 2015; Fudenberg et al. 2016). According to the model, cohesin rings would extrude chromatin until reaching convergent CTCF sites at the *Shh* and at the *Lmbr1* boundary, where they would stall and enable a stable loop to form (Rao et al. 2014; Guo et al. 2015; Vietri Rudan et al. 2015; Wutz et al. 2017)(Rao et al. 2014; Guo et al. 2015; de Wit et al. 2015; Vietri Rudan et al. 2015)(Rao et al. 2014; Guo et al. 2015; de Wit et al. 2015; Vietri Rudan et al. 2015)(Rao et al. 2014; Guo et al. 2015; de Wit et al. 2015; Vietri Rudan et al. 2015)(Rao et al. 2014; Guo et al. 2015; de Wit et al. 2015; Vietri Rudan et al. 2015)(Rao et al. 2014; Guo et al. 2015; de Wit et al. 2015; Vietri Rudan et al. 2015). At the *Lmbr1* boundary, different combinations of CTCF sites can form alternative loops with the i9, i5 and i4 CTCF sites in each cell, but also within the same allele. This is possible, as the CTCF residence time (~1-2 min) on the DNA is shorter than cohesin' s (~20 min) (Hansen et al. 2017), allowing the extruding complex to be dynamically blocked at any of these sites. Accordingly, the derived cHi-C maps represent a static image of all the combination of interactions that exist in individual alleles.

To challenge the 3D structure of the *Shh* domain, the CTCF motifs at the telomeric side were targeted singularly or sequentially in combination. The targeted single i4 and i5 CTCF deletions ranged in size from 11 to 62bp with the aim not to disrupt the *ZRS* enhancer. As expected, the induced mutations led to loss of CTCF binding at the deleted DNA motif, but also resulted in increased binding at neighboring CTCF sites. Similarly, deletion of CTCF sites between the anterior- and posterior-expressing *Hoxa* genes in motor neurons shifts the boundary to the next available CTCF site with the correct orientation (Narendra et al. 2015, 2016). In this configuration, the activating H3K4me3, which normally covers the anterior domain, spreads ectopically to the posterior genes, altering their expression. These studies, together with our data, support the cooperation between CTCF binding sites to maintain functional boundaries.

Although CTCF and cohesin are both necessary for long-range looping, they do have distinct roles: cohesin extrudes the chromatin, while CTCF defines the position of the boundaries (Busslinger et al. 2017; Nora et al. 2017; Rao et al. 2017; Schwarzer et al. 2017; Wutz et al. 2017). As CTCF depletion does not affect the genome-wide loading of cohesin onto the DNA (Nora et al. 2017; Wutz et al. 2017), but its distribution and pausing sites (Busslinger et al. 2017), we asked how the cohesin presence at the *Shh* locus is affected in the CTCF mutants. We observed that deletion of the CTCF sites in the *ΔCTCF i4:i5* mutants led to different distribution of RAD21 at the *Lmbr1* locus. The absence of binding at the deleted CTCF sites was accompanied by increased binding at the ectopic CTCF sites. In accordance with the loop extrusion theory, the cohesin complex, which is unable to be halted at the deleted CTCF i5 in the *ΔCTCF i4:i5* mutants, continues extruding DNA until it reaches

the ZRS CTCF, which is strongly bound by CTCF. In CTCF depleted MEFs, cohesin rings are accumulated at the TSSs of active genes where they are frequently loaded (Busslinger et al. 2017). Here, in the triple deleted *ΔCTCF i4:i5:ZRS* mutants, we did not observe ectopic RAD21 binding or accumulation either at the *Lmbr1* or *Rnf32* TSSs. To assess eventual quantitative changes in RAD21 binding, ChIP-qPCR could be used. However, the absence of prominent effect is likely due to the remaining binding of CTCF at all non-targeted sites of the region, in contrast to the genome wide CTCF depletion in MEFs.

The loss of the i4, i5 and ZRS CTCFs resulted in completely abolished RAD21 binding even from the ectopic Lprom and i9 CTCF sites, although their sequence remained intact. The cooperative loss of cohesin could be the result of decreased CTCF binding strength at these ectopic sites. In fact, it has been suggested that CTCF binding strength is positively correlated with higher probability of loop formation, and thus eventually of cohesin binding (Fudenberg et al. 2016; Clarkson et al. 2019). Here, ChIP-seq for the NIPBL and WAPL factors in limb tissues would reveal where cohesin is loaded and released, respectively, in relation to the CTCF sites of the locus before and after the induced deletions. This information would give a better insight on how localization of the extrusion complex is changed at the locus upon loss of CTCF.

Additionally to the ChIP-seq results, 4C-seq in the single and combinatorial CTCF deletions revealed loss of long-range interactions between *Shh* and the corresponding deleted CTCF sites. Interestingly, the new interactions formed in the mutants correlate with the changes in CTCF binding strength and pattern, as observed by ChIP-seq. Altogether our results support that, although the binding of CTCF and RAD21 to the DNA is independent, there is a dynamic co-dependence between them to achieve a proper 3D structure.

5.2.2 The *Shh*-ZRS enhancer-promoter interaction decreases upon loss of CTCF binding

Genome-wide depletion of CTCF and cohesin experiments result in decreased TAD insulation, disappearance of TADs and loss of loops (Busslinger et al. 2017; Nora et al. 2017; Rao et al. 2017; Schwarzer et al. 2017; Wutz et al. 2017). Here, deletion of two CTCF sites in the *ΔCTCF i4:i5* mutants led to strong loss of the associated loops and to decreased insulation between *Shh* and *Mnx1* TADs. Despite the absence of the anchor loops, the overall TAD structure was weaker but preserved. This is probably due to passive constraints from the intact neighboring CTCF-mediated interactions, within the neighboring *Mnx1* TAD, which constitute a "passive" telomeric limit to the *Shh* TAD. Similarly, in a parallel study at the *Shh* locus, individual CTCF site deletions at the centromeric *Lmbr1* boundary alter the intra-TAD insulation but do not result in TAD fusion (Williamson et al. 2019). Moreover, fusion of the *Sox9* TAD with its neighboring *Kcnj2* TAD is only possible upon deletion of

CTCF sites at the boundary as well as all sites within the TAD as these can be used as backup loop anchors (Despang et al. 2019).

Interestingly, cHi-C in *ΔCTCF i4:i5:ZRS* did not reveal any additional changes in the 3D structure compared to the *ΔCTCF i4:i5*, except the slightly increased de-insulation between *Shh* and *Mnx1* TADs. Based on the ChIP-seq data, the ZRS CTCF binding site is "compensating" the deletion of the centromeric CTCF i5. However, as discussed before the same site does not compensate for the loss of CTCF i4. Therefore, either the ZRS CTCF is not sufficient to dynamically "restore" all lost interactions (Hansen et al. 2017) or it does not bear a buffering role, further underlining the critical importance of CTCFs i4 and i5 for the structure of the locus and the preformed *ZRS-Shh* communication. Indeed, the virtual capture-C derived from both *ΔCTCF i4:i5* and *ΔCTCF i4:i5:ZRS* mutants showed that the *Shh* promoter itself loses the interactions with the telomeric boundary, including the *ZRS*. This interaction profile similarity between the two mutant alleles suggests that, indeed, the ZRS CTCF does not have a compensating function on the structure of the locus.

Further, the polymer-based model of the wildtype locus gave a better visualization of how *Shh* and *ZRS* interact in 3D. More importantly, modeling the *ΔCTCF i4:i5* locus exhibited a significantly increased distance between *Shh* and *ZRS*. Likewise, Williamson et al. showed a decrease in *Shh-ZRS* proximity upon removing one CTCF site by DNA FISH (Williamson et al. 2019). Together, these results suggest that the CTCF loops facilitate the constitutive *Shh-ZRS* interaction.

5.2.3 Loss of CTCF-mediated preformed structure leads to decreased transcription

WISH experiments in the CTCF mutants did not always show evident changes in *Shh* expression pattern (Figure 4. 11). Similarly, Williamson et al. support that the single CTCF deletions, e.g. CTCF 3 which corresponds to the CTCF i5 of our study, did not affect *Shh* expression (Williamson et al. 2019). However, when we used qRT-PCR to quantify expression changes, we found that all single CTCF deletions induced significant reduction in *Shh* transcription, thus questioning the sensitivity of WISH. More importantly, expression analysis in the double mutant *ΔCTCF i4:i5* limb buds revealed a significant 50% decrease in *Shh* transcription. As expected from the similarity of the interaction maps, *Shh* transcription was similarly decreased in *ΔCTCF i4:i5:ZRS* limb buds. Likewise, in genome-wide depletion experiments of CTCF and cohesin, hundreds of genes are misregulated (Nora et al. 2017; Rao et al. 2017; Schwarzer et al. 2017). However, this number of genes is rather small compared to what is expected considering the fundamental role of CTCF and cohesin in chromatin organization and long-range enhancer-promoter communication. When structural variations in mice disturb boundaries with their CTCF sites as well as their large neighboring landscapes, severe gene misregulation is reported due to changes in the overall topology and ectopic enhancer-promoter communication (Lupiáñez et al. 2015). Our results together with these studies highlight

how the characterized as "marginal" expression changes in an *in vitro* cell culture system can be critical *in vivo* for a developing organism.

The expression data in *ΔCTCF i4:i5* and *ΔCTCF i4:i5:ZRS* embryos further indicate that, despite the reduction in *Shh-ZRS* distance, their communication in the 3D space of the nucleus can still be achieved. The remaining 50% of *Shh* expression is possibly mediated by CTCF-independent mechanism(s). Recently, the ubiquitously expressed YY1 protein has been proposed to mediate enhancer-promoter interactions (Beagan et al. 2017; Weintraub et al. 2017). YY1 is strongly enriched in genome-wide loops in human cell lines (Rao et al. 2014), precisely binding at active enhancers and promoter proximal elements (Beagan et al. 2017; Weintraub et al. 2017). In particular, genome-wide CTCF occupancy decreases during differentiation from ESCs to NPCs, while YY1 occupancy increases at NPC-specific enhancer-promoter interactions at these loci (Beagan et al. 2017). Accordingly, knockdown of YY1 in NPCs results in interruption of the NPC-specific enhancer-promoter looping. Similarly, CRISPR-Cas9 targeting of the YY1 motif in the regulatory region of the *Raf1* and *Etv4* genes in mESCs resulted in decreased enhancer-promoter contact frequency and mRNA levels (Weintraub et al. 2017). Targeted recruitment of a dCas9-YY1 fusion protein to a mutated YY1 binding site at the promoter-proximal region of *Etv4* restores the looping with its target enhancer and increases its transcriptional level. Interestingly, Beagan et al. reported that YY1-mediated looping interactions in NPCs are formed frequently adjacent to and nested within constitutive CTCF-CTCF loops (Beagan et al. 2017). Moreover, an inter-species analysis of CTCF binding showed that one third of CTCF-bound regions in primate lymphoblastoid cell lines are co-bound by YY1, they are associated with increased transcriptional activity and contribute to each other's evolutionary stability (Schwalie et al. 2013). It is thus tempting to test if there are YY1 binding sites within the *Lmbr1* gene body and whether this could be the backup mechanism maintaining the residual *Shh-ZRS* interaction in the *Shh*-expressing limb cells of *ΔCTCF i4:i5* and *ΔCTCF i4:i5:ZRS* mutants. Another possibility is that the *Shh-ZRS* communication is achieved through alternative looping factors either tissue-specific or ubiquitous, e.g. ZNF143, Mediator (Phillips-Cremins et al. 2013; Heidari et al. 2014; Rao et al. 2014; Bailey et al. 2015). Mediator, for instance, interacts with cohesin in a CTCF-independent manner (Phillips-Cremins et al. 2013), but it is not known if it can mediate such long-distance enhancer-promoter interactions. Accordingly, the chromatin folding of the *Shh* locus in the absence of the CTCF-mediated loop probably relies on the presence of other DNA-binding proteins that act to bridge, *via* an unknown mechanism, the long-range interaction of the *ZRS* enhancer with its target promoter.

5.3 *Lmbr1* constitutive transcription has an auxiliary role in chromatin structure and *Shh* regulation

Along with the discovery of TADs, it was observed that boundaries associate strongly with housekeeping genes (Dixon et al. 2012; Rao et al. 2014). In many other species that demonstrate domain-like chromatin organization in the nucleus, transcription seems to be the driving force behind it. In *Drosophila*, TAD borders are mainly associated with actively transcribed gene-dense regions (Szabo et al. 2019). Inhibition of either transcription or transcription initiation induces decrease in domain organization (Rowley et al. 2017). Additionally, paused RNA Polymerase was suggested to mediate preformed enhancer-promoter interactions during *Drosophila* embryogenesis (Ghavi-Helm et al. 2014). In vertebrates, although there is a lot of evidence toward the involvement of transcription in organization of the genome, how determinant is its role and the exact mechanism have not yet been elucidated. At the *Shh* TAD, the telomeric boundary is also occupied by the constitutively expressed gene, *Lmbr1*. Therefore, we tested *in vivo* the contribution of its transcription to the preformed configuration and *Shh* regulation.

5.3.1 Abrogation of *Lmbr1* transcription redistributes the chromatin interactions at the *Shh* locus

Ultra-deep Hi-C revealed that some TAD boundaries are depleted from CTCF binding, but enriched in active gene promoters and their associated active histone marks (Bonev et al. 2017). For instance, during differentiation from mESCs to NPCs, a new boundary forms at the TSS of the *Zfp608* neural gene together with its transcriptional activation, but without the existence of a CTCF binding site. However, induced gene expression does not create a new boundary, suggesting that transcription *per se* is not sufficient to form the boundary (Bonev et al. 2017). Here, the telomeric boundary of the *Shh* TAD is occupied by both CTCF and the *Lmbr1* housekeeping gene. CHi-C performed in *Lmbr1* promoter deficient mouse embryos did not disrupt the TAD boundary, but did show a slightly changed interaction pattern within *Lmbr1.*

In the transcriptionally inactive mouse sperm, promoters and enhancers are enriched with active histone marks, e.g. H3K4me3, H3K27ac, H3K4me1, while boundaries are occupied by CTCF and cohesin (Jung et al. 2017). These properties are sufficient for the formation of TAD-like domains resembling the structure in ESCs or fibroblasts (Battulin et al. 2015; Jung et al. 2017). Thus, the chromatin organization of the sperm genome is not dependent on transcription *per se* but transcription-associated factors cooperating with architectural proteins (Jung et al. 2017). Interestingly, in the *Lprom* mice, 4C-seq experiments demonstrated that increased contacts were formed between the *Shh* promoter and the CTCF sites i5 and i9 centromeric to the *ZRS*. Hence,

abrogation of the *Lmbr1* transcription induced redistribution of the CTCF/cohesin-dependent chromatin interactions. Supporting this finding, several studies have shown an interconnection between architectural proteins and transcription but through different mechanisms. According to Busslinger et al., transcription is necessary for the correct positioning of cohesin molecules on the DNA. Cohesin is loaded by the NIPBL loading factor often at active TSSs and translocates along the DNA in a transcription-dependent manner accumulating at CTCF sites (Busslinger et al. 2017). An alternative trigger for the translocation of cohesin along the DNA fiber is the negative supercoiling, which occurs simultaneously with ongoing transcription (Racko et al. 2018). Conversely, elongating Pol II can also displace cohesin, which alters the locus compaction and the 3D organization (Heinz et al. 2018). Further experimentation on the *Lprom* mutants concerning the changes in CTCF and cohesin binding at the locus additionally to high-resolution microscopy could give insight into the mechanism that transcription induces local chromatin conformation changes.

Despite the contrasting views, there is an intricate relation between architectural proteins and transcription which supports its contribution in 3D chromatin interactions. Whether it is transcription or transcription-associated factors that have more central role needs to be further investigated.

5.3.2 Loss of constitutive transcription has an impact on *Shh* transcription efficiency

In *Drosophila*, inhibition of transcription, transcription initiation and/or transcription elongation genome-wide has resulted in alterations of TAD formation (van Steensel and Furlong 2019). To date, the results in mouse are inconclusive about the role of constitutive transcription in inducing looping and TAD formation, and even more elusive about its role on enabling regulatory interaction between enhancers and promoters (Giorgetti et al. 2016; Bonev et al. 2017; Ke et al. 2017).

At the *Shh* locus, the preformed structure which brings together *Shh* and the *ZRS* enhancer constitutes a great opportunity to assess the loss of transcription not only in relation to TAD formation, but also to the long-range interaction between *Shh* and the *ZRS*. Indeed, deletion of the *Lmbr1* promoter led to a 20% loss of *Shh* transcription. Assessing the enhancer accessibility and activity did not show evidence of a less potent ZRS that could be the cause of the mild *Shh* expression loss. However, 4C-seq in the *Lprom* mutants revealed a centromeric shift of interactions within the *Lmbr1* gene along with a loss in *Shh-ZRS* contact. This conformational change is likely less favorable to the *Shh-ZRS* communication and *Shh* transcription.

Furthermore, it has been suggested that preformed topologies favor the creation of transcriptional hubs, where increased polymerase and co-factors facilitate transcription robustness (Furlong and Levine 2018). This process could be mediated *via* phase separation (Hnisz et al. 2017). According to this model, molecules like TFs, transcription co-activators, Pol II, undergo chemical modifications

that allow for multicomponent interactions by forming intermolecular bonds. These bonds function as "crosslinks" and confer higher stability in their interactions. Live-cell visualization has captured Mediator and Polymerase forming such phase-separated transcriptional condensates in the nucleus (Cho et al. 2018). It is possible that the constitutively transcribed *Lmbr1,* embedded in a preformed topology, brings the transcriptional apparatus in place and facilitates its stabilization by favoring stable bonds between transcription-associated factors. Therefore, disruption of transcription would probably create a less favorable for transcription environment and would result in decreased transcription robustness.

Overall, preformed structures reduce the search in space and time of an enhancer to find its target promoter. It seems plausible that the involvement of a housekeeping gene in this topology reduces further the time for transcription initiation. As for now, the underlying mechanism that active transcription acts on the 3D architecture is still to be studied. However, our data indicate that although transcription does not seem to be the primary driver in structuring the genome, it is still a prominent factor influencing it. In summary, our *in vivo* data indicate a multicomponent-based *Shh* locus organization, where transcription has a complementary role to CTCF to maximize *Shh* transcription levels.

5.4 The preformed topology ensures *Shh* transcriptional robustness for normal limb development

Our study shows that perturbation of either the *Lmbr1* transcription or CTCF binding results in altered chromatin topology and decreased *Shh* transcription, without resulting in skeletal phenotypes. Moreover, targeted alterations of the regulatory landscape and of the gene itself have shown that a specific transcriptional threshold needs to be reached to induce loss-of-function phenotypes (Matsumaru et al. 2011; Lettice et al. 2014; Symmons et al. 2016). Specifically, heterozygous *Shh* animals do not exhibit any phenotype, suggesting that a remaining 50% of transcription is enough to develop normal digits. This genetic phenomenon is also seen at other development loci such as *Hoxd13*, where only homozygote animals show strong phenotype including digit size reduction and metacarpal defects (Dollé et al. 1993). Moreover, it was reported that residual *Shh* expression levels as low as 10% can produce a normal number of digits but with other digit-related phenotypes (Lettice 2014). This robustness in *Shh* expression makes it difficult to assess the potential effect of any partial perturbation in *Shh* regulation.

In support of the above observations, transcription is characterized by fluctuations, or else cell-to-cell heterogeneity in the mRNA output, but which under normal genetic or environmental conditions are not harmful (Hobert 2010). When a component is removed or disrupted, the system

becomes less resistant to fluctuations and leads to mutant phenotypes. Similarly, here, the CTCF mutants induce up to 53% loss of *Shh* transcription, which does not lead to a limb phenotype. This is reminiscent of a study in the mouse limb, where one copy of *Gli3* or *Shox2* does not cause any phenotype (Osterwalder et al. 2018). Compound deletions, further decreasing the gene transcription below 30%, were required to induce limb malformations. Similarly, at the *HoxD* locus, digit enhancer deletions in a heterozygote *Hoxd13* background leads to digit malformations (Montavon et al. 2011). Accordingly, we further perturbed the *ΔCTCF i4:i5 Shh* locus to assess the effect of the disrupted preformed topology in adverse genetic conditions.

5.4.1 The integrity of the *ZRS* enhancer is essential for *Shh* regulation

The centromeric part of the *ZRS* was proposed to mediate the essential long range interaction with *Shh* promoter (Lettice et al. 2014). Transgenic mouse limbs carrying the enhancer and a *LacZ* reporter gene instead of the centromeric "long-range regulatory" *ZRS* module exhibited normal posterior-distal LACZ staining, suggesting that the telomeric part of the enhancer is sufficient to determine the spatiotemporal expression of *Shh* (Figure 1. 3). However, these mice could not activate *Shh* resulting in digit agenesis in the forelimbs and showed variable degree of oligodactyly in the hindlimbs. Therefore, the missing part of the enhancer was suggested to be responsible for the long-range interaction of the *ZRS* enhancer with *Shh* (Lettice et al. 2014).

Here, the *ΔZRSreg* allele, which bears approximately the same genomic deletion, does not affect the overall *Shh* TAD or the anchor loops. Despite a slightly higher interaction with neighboring CTCF sites, the contact with the *ZRS* remains unaffected, suggesting that this *ZRS* module does not have impact on the preformed topology at the *Shh* locus and does not participate in its establishment. In support of this, 3D DNA FISH showed that complete deletion of the enhancer sequence does not alter the proximity between *Shh* and the *ZRS* region (Amano et al. 2009). However, expression analysis in *ΔZRSreg* forelimbs and hindlimbs showed a 75% loss of *Shh* transcription, which led to an oligodactyly phenotype in the forelimbs, but to normal hindlimbs. This phenotype resembles the one acquired by Lettice et al., although the severity differs probably due to experimental differences either in the breakpoints of the genomic deletion or in the mouse genetic background (Lettice et al. 2014). Moreover, the phenotypical difference between forelimbs and hindlimbs has been already described in detail (Chiang et al. 2001), but its molecular cause has not yet been understood. Together these results suggest that the centromeric part of the *ZRS* enhancer includes TFBSs which are important for normal *Shh* regulation.

In fact, several ETS sites (ETS2, 3, 4 and 5) are located in the centromeric part of the enhancer and thus, deleted in the *ΔZRSreg* allele (Figure 4. 12). Transgenic analysis of mouse embryos carrying *ZRS* mutations removing different number of ETS sites showed that, in the absence of the ETS2-5,

Shh expression domain is smaller than in wildtype embryos (Lettice et al. 2012). Thus, the whole *ZRS* sequence is necessary for proper quantitative and spatiotemporal *Shh* expression. Except for the ETS sites, the ZRS CTCF site is also found at the deleted centromeric end of the *ZRS* enhancer (Figure 4. 12). The *ΔCTCF i4:i5:ZRS* experiments showed that in the absence of CTCF i4 and i5, the ZRS CTCF is not able to compensate for their loss in the 3D conformation of the locus. Yet this does not exclude a potential role of the ZRS CTCF in the intact preformed structure of the locus. Similarly, if other architectural or looping proteins associate with this part of the *ZRS* in order to mediate *Shh-ZRS* colocalization in the posterior limb, their absence would affect the efficiency of *Shh* transcription. Therefore, the transcriptional output of the *ΔZRSreg* originates primarily from the absence of binding sites for trans-acting factors and possibly from changes in local chromatin interaction.

5.4.2 Disruption of the preformed structure in a sensitized genetic background leads to *Shh* loss-of-function phenotype

To assess the perturbation of the preformed structure in a sensitized genetic background, the *ΔZRSreg* mutation, which creates a "compromised" enhancer, was engineered in *ΔCTCF i4:i5* ESCs. In E10.5 limb buds, cHi-C in *ΔCTCF i4:i5: ΔZRSreg* revealed a 3D structure, which largely resembled the *ΔCTCF i4:i5* and *ΔCTCF i4:i5:ZRS* structures. By comparing the virtual capture-C profiles of *ΔCTCF i4:i5, ΔCTCF i4:i5:ZRS* and *ΔCTCF i4:i5: ΔZRSreg,* we showed similar loss of interactions between the *Shh* region and the *Lmbr1* gene body, but did not observe additional decrease in *Shh-ZRS* contact frequency in the latter mutant. Nevertheless, the absence of the preformed conformation in the hypomorphic allele led to the loss of *Shh* transcription, fully penetrant digit agenesis in the forelimbs and oligodactyly in the hindlimbs. Accordingly, as the sole loss of CTCF does not reduce transcription below 50%, its association with a hypomorphic *ZRS* proves its importance to buffer genetic alterations. Similarly, Hi-C in mouse Th2 cells shows that, intra-TAD CTCF binding associated with active enhancers acts to stabilize enhancer-gene promoter contacts (Ren et al. 2017). Upon knockdown of CTCF and specific CTCF motif targeting no striking gene expression changes are observed. However, flow cytometry experiments revealed significant cell-to-cell gene expression variability, indicating that CTCF ensures robustness in gene expression levels by stabilizing long-range enhancer-promoter interactions (Ren et al. 2017). Accordingly, we suggest that the CTCF-mediated preformed structure confers robustness and ensures maximal *Shh* expression levels.

The role of CTCF in transcriptional robustness gives an alternative explanation to the minimal gene expression changes observed in genome-wide CTCF depletion experiments (Nora et al. 2017). Yet, even minimal gene expression differences can be critical during development, differentiation and evolution (Barroso et al. 2018; Urban and Johnston 2018). For instance, fluctuations in *Nanog*

expression in mESCs define cell fate decisions (Kalmar et al. 2009). Such fluctuations have been shown to originate from alterations either in the amount of transcribed mRNA or in the frequency at which transcription initiates at a promoter (Bartman et al. 2016; Larsson et al. 2019). Enhancers, but also the chromatin architecture they are embedded in, can cause such transcriptional fluctuations by influencing bursting frequency in order to foster it or to buffer it. Deletion of the *β-globin* enhancer LCR during erythroid maturation dramatically decreases the frequency of bursts with marginal effects on the amount of mRNA molecules produced (Bartman et al. 2016). Conversely, forced LCR- *β-globin* promoter looping increases strongly the frequency of transcriptional bursts. At the *Shh* locus, our data suggest that the preformed structure ensures maximal transcription by enabling the stable interaction of *Shh* with its *ZRS* enhancer.

Considering that SHH is an important morphogen, the "excess" of *Shh* transcription might result in the minimal expression level ensuring that all cells within the ZPA express *Shh* at a certain amount and at the right timepoint. Indeed, it is suggested that expression of SHH occurs in temporal waves and is generally accepted that sustained SHH signaling is necessary for limb development (Tickle and Towers 2017; Zhu and Mackem 2017). Presumably, this is due to the involvement of SHH not only in antero-posterior patterning, but also in limb outgrowth *via* feedback circuits. Hence, it is intriguing to hypothesize whether and how decreased transcriptional robustness impacts differentially the cell populations of limb buds. In fact, the various phenotypes of this study could support the hypothesis that the loss of *Shh* transcription does not occur homogeneously throughout the limb. Instead, due to increased transcription fluctuations, the loss is heterogeneous and appears as an averaged decrease. This would probably alter the concentration gradient of SHH, and consequently the development of all skeletal segments of the limb including the digits. To address this question experimentally, single cell RNA-seq experiments in wildtype and in the mutant limbs would be an appropriate strategy.

5.5 Impact of preformed topologies on gene regulation

One of the important questions regarding enhancer-promoter communication in the nucleus is whether it pre-exists in precursor cells (permissive) or whether it forms dynamically along with gene activation (instructive) (de Laat and Duboule 2013). The existence of both cases means that proximity does not automatically translate into transcriptional activation, but also that sometimes it might not be even necessary. For instance, it has been suggested that *Sox2* transcription in living ESCs occurs without evident increase in enhancer-promoter proximity (Alexander et al. 2018). On the other hand, live imaging of enhancer-promoter interactions and transcription simultaneously

in *Drosophila* showed that physical proximity is essential for initiation and maintenance of transcription and its sustainability relates to higher transcription levels (Chen et al. 2018).

Although the two modes of enhancer-promoter communication seem not to overlap, in some loci they do coexist (Freire-Pritchett et al. 2017). During mESC differentiation, most enhancer-promoter contacts occurred concomitantly with gene activation, but there were few loci identified with pre-established contacts (Bonev et al. 2017). Moreover, 4C-seq experiments at the *HoxD* locus showed that *Hoxd13* forms constitutive interactions with regulatory elements within the centromeric TAD independently of its transcriptional status, while *Hoxd9* and *Hoxd11* establish expression-dependent contacts with CREs (Andrey et al. 2013). At the *Pitx1* locus, a dynamic chromatin conformation ensures that a fore- and hindlimb enhancer will activate *Pitx1* only in the hindlimb (Kragesteen et al. 2018). Thus, instructive dynamic interactions can refine unspecific enhancer activities into more specific transcriptional output, even when they coexist with preformed contacts. In keeping with this, alteration of these interactions can cause gene misregulation. For instance, forced chromatin looping between the *β-globin* promoter and its *LCR* enhancer in primary erythroid progenitor cells induces ectopic *β-globin* expression (Deng et al. 2012).

Conversely, permissive interactions, that potentially reduce the search time and space of enhancers to find their target promoter, increase transcription efficiency. Preformed enhancer-promoter contacts have been noted in various loci and cell types mediated by Pol II (Ghavi-Helm et al. 2014), PRC2 (Cruz-Molina et al. 2017), cohesin (Rubin et al. 2017b). At the *Shh* locus this preformed interaction, nested in the *Shh* TAD boundaries, is mediated by CTCF. In contrast to dynamic interactions, the constant proximity of the *ZRS* to *Shh* does not restrict its activity, but promotes *Shh* transcriptional robustness (Figure 5. 1). Alternatively, transcriptional robustness can be ensured *via* redundancy of regulatory elements. Indeed, studies in *Drosophila* have shown that the addition of multiple enhancer activities (Hong et al. 2008; Frankel et al. 2010) or multiple promoter activation by a single enhancer (Fukaya et al. 2016) can account for such transcriptional robustness. Similarly, in mouse, nine enhancer elements with additive and partially redundant activities compose the final expression pattern of *Ihh* in limbs and skull (Will et al. 2017). Therefore, genetic alterations affecting either the composition of the enhancer cluster or the gene dosage result in very diverse phenotypical outcomes. Furthermore, an extensive study on redundant limb enhancers showed that a lot of developmental genes are regulated by multiple enhancers (Osterwalder et al. 2018). Deletion of single enhancer elements was not sufficient to drive gene expression changes, while deletions of the same enhancers in a sensitized genetic background, e.g. one copy of the target gene, resulted in severe limb malformations, supporting that enhancer redundancy confers transcriptional and phenotypic robustness. In contrast to these loci, knock-out

of the *ZRS* leads to complete *Shh* loss of function in the limb (Figure 4. 17, Sagai et al. 2005). It is thus critical to secure robust *Shh-ZRS* communication for normal *Shh* expression in the limb and this is achieved through the special 3D structural configuration of the locus.

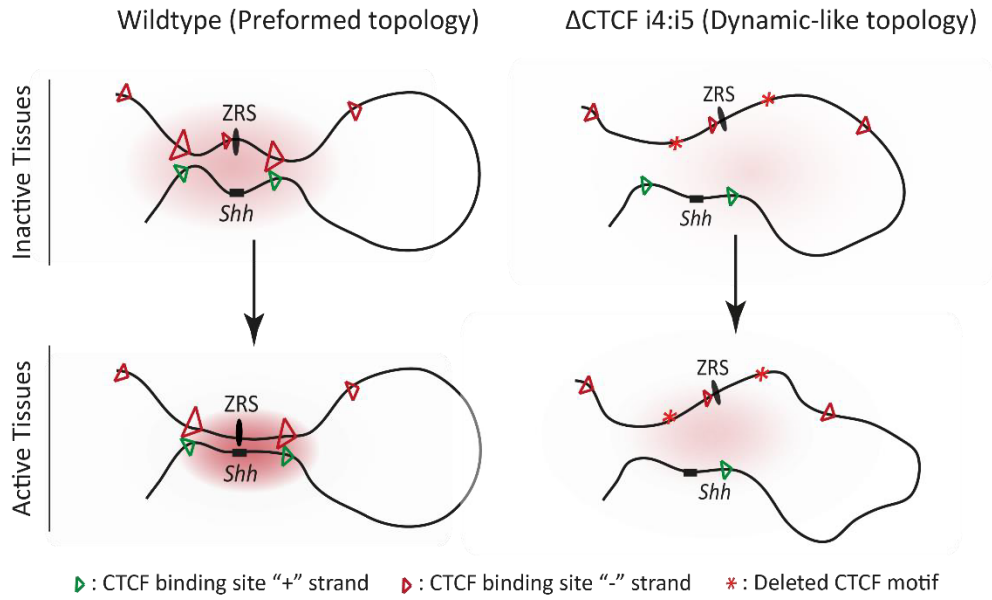


Figure 5. 1 Model of the preformed topology at the Shh l

The preformed topology at the *Shh* locus is mediated by CTCF sites, by forming a loop between *Shh* and its limb enhancer, *ZRS* and thus, bringing them in close proximity even in inactive tissues. In active tissues, *Shh-ZRS* colocalization (Williamson et al. 2016) is increased and *Shh* is expressed. Conversely, in the *ΔCTCF i4:i5* mutants, where the CTCF sites are deleted, the distance between *Shh* and the *ZRS* is increased, reminiscent of dynamic topologies. Upon TF binding, *ZRS* finds and activates *Shh* less efficiently. Therefore, preformed topologies create an "environment" ready for gene activation which leads to transcriptional robustness.

5.6 The preformed topology at the *Shh* locus in respect to disease and evolution

Mutations in regulatory elements are cause of disease, but also drivers of evolutionary change. Numerous single-nucleotide substitutions and duplications of the *ZRS*, resulting in *Shh* misexpression in the anterior limb, have been reported in human patients with limb malformations (Wieczorek et al. 2010; Klopocki and Mundlos 2011; Al-Qattan 2018). One could hypothesize that this misexpression is aided by the preformed *Shh-ZRS* loop. It is possible that recruitment of non-tissue-specific TFs to the *ZRS* can activate *Shh* ectopically due to their stable interaction even in *Shh* non-expressing cells. Moreover, patients with a condition known as Acheiropody (OMIM #200500) have bilaterally amputated-like extremities with absence of hands and feet, resembling *Shh* loss-of-function phenotype (Ianakiev et al. 2001). These patients bear a deletion spanning *Lmbr1* exon 4 and parts of introns 3 and 4. Surprisingly, the *ZRS* is not included in the deletion. However, the

deleted sequence has in the proximity CTCF binding sites which loss could destabilize the 3D structure at the human *SHH* locus. To better understand the molecular basis of this condition, exact characterization of the deletion breakpoints and sequencing of the *ZRS* enhancer to exclude any further mutations are primarily necessary.

Admittedly, the *ZRS* is vulnerable to accumulating mutations, as any change would affect only the development of the limb. For instance, deletion of ETS and HOXD13 binding sites leads to loss of the *ZRS* enhancer function and ultimately, failure in limb outgrowth in pythons (Kvon et al. 2016; Leal and Cohn 2016). However, both the *ZRS* sequence and the *Shh* regulatory domain are well conserved among most vertebrates. It is interesting that the *Lmbr1* locus was integrated in the regulatory landscape of *Shh* within the vertebrate lineage upon a genomic rearrangement (Irimia et al. 2012). This finding raises the question whether *Lmbr1* brought into the *Shh* regulatory domain the sequences involved in the preformed structure or if they evolved later. Although the sequence of events is not known, the high conservation in vertebrates underlines the significance of this genomic structure, which is then mirrored in the regulatory structure. Overall, preformed topologies can be seen as a friendly "environment" for enhancer evolution, as they offer an already formed structure and ensure the transcriptional outcome (de Laat and Duboule 2013).

5.7 Conclusion and Outlook

The focus of this study was to shed light on the formation of the preformed topology of the *Shh* locus and on its role in *Shh* expression in the limb. I first characterized the regulatory domain of *Shh* to define candidates that mediate the long-range stable interaction with its sole limb enhancer, the *ZRS*. Then, I genetically perturbed individually these candidates and analyzed the effect on *Shh* transcription and further, on digit development of mouse embryos.

This is the first locus-specific study which genetically dissects both the role of CTCF and transcription in respect to gene regulation and chromatin architecture. Our findings specifically,

1. support the involvement of constitutive transcription in the overall 3D organization of the locus,
2. present a dynamic interplay between CTCF sites and cooperation with the cohesin complex and,
3. explore the significance of the CTCF-mediated preformed topology in *Shh* regulation by ensuring transcriptional robustness.

However, there are still open questions and mechanisms to be further investigated. In what way transcription impacts the 3D chromatin organization of the locus? Does it have an additive or a parallel role in the preformed topology? What is the CTCF-independent mechanism which supports the maintenance of the long-range *Shh-ZRS* interaction? How the reduced trancriptional robustness

affects *Shh* secretion? Do all cells of the ZPA express less SHH or is there more expression variability between cells? Was the *ZRS* sequence acquired together with the integration of the *Lmbr1* at the *Shh* regulatory domain or it evolved later? Combinatorial deletion of the *Lmbr1* promoter with the CTCF sites would answer if the role of transcription is additive. Further characterization of the *ZRS* sequence would probably unravel more trans-acting factors potentially involved in loop formation. Moreover, visualization in space and time by high resolution microscopy in combination with structural data would be an exciting approach to elucidate the exact mechanisms behind *Shh* regulation in the nucleus. However, it would be challenging to apply it in mouse embryos.

In a broader perspective, our study shows how important it is to assess the regulation of developmental genes in the right tissue and time of expression. More importantly, it underlines that transcriptional robustness confers fitness in normal and adverse genetic or environmental conditions and that in some cases, e.g. developmental genes, might be more critical than in others. Finally, as not so many developmental genes are regulated by a single enhancer, it is appealing to further investigate and reveal the molecular mechanisms behind the organization of this locus and the *ZRS* regulation on *Shh.*

6 SUMMARY

During embryogenesis, long-range gene regulation involving physical proximity between enhancers and promoters generates precise patterns of gene expression in space and time. These functional chromatin interactions are thought to be constrained by domains of preferential interactions called topologically associating domains (TADs). In some cases, physical proximity between enhancers and promoters coincides with gene activation, whereas, in others, preformed topologies already exist before gene transcription. The *Shh* locus constitutes a paradigm for the latter. The only limb enhancer, the *ZRS,* is located 1 Mb away from the *Shh* gene promoter, which is embedded within a TAD boundary and the constitutively expressed *Lmbr1* gene. However, the factors supporting this preformed topology and their role in *Shh* regulation have not been elucidated yet.

In this study, we set out to investigate the preformed configuration underlying *Shh* regulation by the *ZRS* enhancer *in vivo* during mouse embryogenesis. Using capture Hi-C, we assessed the chromatin structure of the *Shh* locus in E10.5 mouse embryonic limb buds. We found that the *Shh* TAD is characterized by a strong CTCF-associated loop interaction, linking the *Shh* region embedded in the centromeric TAD boundary to the *Lmbr1* gene region embedded in the telomeric boundary and containing the *ZRS* enhancer. Interestingly, the same contacts were also observed in mESCs and E10.5 mouse midbrain independently of the gene or the enhancer activity, thus confirming that the *Shh* locus adopts a preformed configuration and that the *Shh-ZRS* proximity is tissue-invariant. To test the role of the constitutive transcription in establishing this preformed configuration, we deleted the *Lmbr1* promoter by CRISPR-Cas9. Loss of transcription led to a shift of the *Shh* interactions toward the centromeric *Lmbr1* gene body, a slightly reduced *Shh-ZRS* contact frequency and a 20% reduction in *Shh* transcription. Next, we targeted the binding motifs of the most prominent CTCF sites surrounding the *ZRS* and underlying the looping events observed in the cHi-C maps. Single or combinatorial CTCF site deletions revealed a cooperative CTCF and cohesin binding to maintain the boundary functional. Moreover, loss of CTCF around the *ZRS* resulted in disruption of the *Shh–ZRS* preformed interaction and a 50% decrease in *Shh* expression. Interestingly, this decrease in *Shh* transcription was not sufficient to lead to a skeletal phenotype, suggesting that the residual enhancer-promoter communication occurs *via* an additional, CTCF-independent mechanism. However, the combinatorial loss of CTCF binding along with a hypomorphic *ZRS* allele resulted in almost complete loss of *Shh* transcription and thus, severe *Shh* loss of function and digit agenesis.

In summary, our results demonstrate that the preformed chromatin structure of the *Shh* locus is sustained by multiple components and acts to reinforce enhancer–promoter communication in

order to ensure transcriptional robustness. Moreover, our data suggest that constitutive transcription at boundaries has a secondary role in chromatin organization, which is mainly driven by CTCF and the cohesin complex. Overall, our study underlines that the implications of destabilized chromatin organization for gene regulation need to be studied at endogenous genomic context, at the right time and in the right tissue, and in particular when complex morphogenetic processes are concerned.

7 ZUSAMMENFASSUNG

Die physikalische Interaktion von Chromatin zwischen Genpromotoren und regulatorischen Elementen (Enhancer) im Zellkern hat einen entscheidenen Einfluss auf die räumliche und zeitliche Genexpression während der Embryogenese. Untersuchungen der Chromatinstruktur in Wirbeltieren haben zur Entdeckung von sogenannten "Topologically associating domains" (TADs) geführt. TADs sind Bereiche im Genom die eine hohe Interaktionfrequenz von Chromatin aufweisen und dazu beitragen, dass Enhancer zu ihren Zielgenen finden. Abhängig von der Aktivität der Zielgene lassen sich zwei Typen von Chromatinstrukturen unterscheiden. Zum einen führen dynamische Chromatin-Interaktionen von Promotoren und Enhancern unmittelbar zur Genexpression, wohingegen in anderen Fällen stabile Interaktionen bereits vor der Genexpression existieren. Im Falle des *Shh* Gens liegt ein solche stabile Chromatin-Interaktion mit dem *ZRS* Enhancer vor. Der *ZRS* Enhancer, der die Expression in den sich entwickelnden Extremitäten reguliert, liegt ungefähr eine Megabase von dem *Shh* Promoter entfernt, innerhalb einer TAD-Grenze und des konstitutiv exprimierten *Lmbr1* Gens. Die Faktoren, die zu dieser stabilen Interaktion führen und welche Rolle sie bei der Regulation von *Shh* spielen sind jedoch nur unzureichend verstanden.

In der vorliegenden Arbeit wurde die Rolle der stabilen Interaktion zwischen *Shh* und dem *ZRS* Enhancer während der Embryonalentwicklung in der Maus untersucht. Die Untersuchung der Chromatinstruktur in den sich entwickelnden Extremitäten mittels capture Hi-C zeigten spezifische Interaktion von CTCF-gebundenen Chromatinbereichen am *Shh* Lokus. Diese konnten als Chromatinschleifen oder Loop-Interaktionen identifiziert werden. Diese Loop-Interaktionen verbinden die in der zentromeren TAD-Grenze eingebettete *Shh* Region mit der telomeren TAD-Grenze, welche *Lmbr1* und den *ZRS*-Enhancer enthält. Interessanterwiese handelt es sich dabei um stabile Chromatinstrukturen, die unabhängig von der *Shh* Genaktivität auftreten. Die gleichen Interaktionen konnten auch in embryonalen Stammzellen und Gehirngeweben der Maus identifiziert werden. Um die Rolle der konstitutiven Transkription von *Lmbr1* beim Aufbau dieser stabilen Konfiguration zu testen, wurde der *Lmbr1* Promotor mit Hilfe des CRISPR-Cas9 Systems entfernt. Der Transkriptionverlust führte zu einer Verschiebung der *Shh*-Interaktionen in Richtung des zentromeren *Lmbr1* Genkörpers, einer leicht verringerten *Shh*-*ZRS*-Interaktionsfrequenz und einer Verringerung der *Shh* Transkription um 20%. Als Nächstes wurden die CTCF-Bindungsmotive in der Umgebung des *ZRS* Enhancers durch Nutzung des CRISPR-Cas9 Systems mutiert. Die Deletion einzelner Bindungsmotive, sowie deren kombinierte Deletionen, führt zum Verlust des CTCF-Bindefaktors und eines weiteren DNA-Bindefaktors, Cohesin. Zudem wurden dynamischen Veränderungen an benachbarten Bindungsmotiven beider Faktoren gemessen, die zur Stabilität

der beobachteten Loop-Interaktionen beitragen. In den CTCF-Deletionmutanten konnte eine deutliche Abnahme in der Loop-Interaktionfrequenz und eine 50%ige Abnahme der *Shh* Expression während der Extremitätenentwicklung gemessen werden. Dieser Expressionsverlust hatte jedoch keine phänotypischen Auswirkungen in den Deletionsmutanten, was darauf hindeutet, dass die restliche Enhancer-Promotor-Interaktion über einen zusätzlichen, CTCF-unabhängigen Mechanismus erfolgt. Erst die kombinierte Deletion von CTCF-Bindungsmotiven und einem hypomorphen *ZRS*-Allel führte zu einem fast vollständigen Verlust der *Shh* Expression und damit zu einem schweren *Shh* Funktionsverlust und Gliedmaßen-Agenesie.

Zusammenfassend zeigen die hier präsentierten Ergebnisse, dass die stabile Chromatinstruktur am *Shh* Lokus durch mehrere Komponenten aufrechterhalten wird und die Enhancer-Promotor-Interaktion verstärkt, um eine robuste Transkription während der Embryonalentwicklung sicherzustellen. Darüber hinaus legen die Daten nahe, dass konstitutive Transkription an TAD-Grenzen eine untergeordnete Rolle bei der Chromatinorganisation spielen. Stabile Chromatinstrukturen werden hauptsächlich durch CTCF und den Cohesin-Komplex determiniert. Insgesamt unterstreicht die Studie, dass, insbesondere wenn komplexe morphogenetische Prozesse betroffen sind, Auswirkungen einer destabilisierten Chromatinorganisation auf die Genregulation im endogenen genomischen Kontext und zum richtigen Zeitpunkt in der Embryonalentwicklung untersucht werden müssen.

8 REFERENCES

Adli M. 2018. The CRISPR tool kit for genome editing and beyond. *Nat Commun* **9**: 1911.

Ahituv N, Zhu Y, Visel A, Holt A, Afzal V, Pennacchio LA, Rubin EM. 2007. Deletion of Ultraconserved Elements Yields Viable Mice ed. R.A. Gibbs. *PLoS Biol* **5**: e234.

Akhtar W, de Jong J, Pindyurin A V, Pagie L, Meuleman W, de Ridder J, Berns A, Wessels LF a, van Lohuizen M, van Steensel B. 2013. Chromatin position effects assayed by thousands of reporters integrated in parallel. *Cell* **154**: 914–27.

Al-Qattan MM. 2018. Zone of Polarizing Activity Regulatory Sequence Mutations/Duplications with Preaxial Polydactyly and Longitudinal Preaxial Ray Deficiency in the Phenotype: A Review of Human Cases, Animal Models, and Insights Regarding the Pathogenesis. *Biomed Res Int* **2018**: 1–9.

Alexander JM, Guan J, Huang B, Lomvardas S, Weiner OD. 2018. Live-Cell Imaging Reveals Enhancer-dependent Sox2 Transcription in the Absence of Enhancer Proximity. *bioRxiv* 409672.

Amano T, Sagai T, Seki R, Shiroishi T. 2017. Two Types of Etiological Mutation in the Limb-Specific Enhancer of Shh. *G3 (Bethesda)* **7**: 2991–2998.

Amano T, Sagai T, Tanabe H, Mizushina Y, Nakazawa H, Shiroishi T. 2009. Chromosomal Dynamics at the Shh Locus: Limb Bud-Specific Differential Regulation of Competence and Active Transcription. *Dev Cell* **16**: 47–57.

Anderson E, Devenney PS, Hill RE, Lettice L a. 2014. Mapping the Shh long-range regulatory domain. *Development* **707**: 108480.

Anderson E, Hill RE. 2014. Long range regulation of the sonic hedgehog gene. *Curr Opin Genet Dev* **27**: 54–59.

Anderson E, Peluso S, Lettice LA, Hill RE. 2012. Human limb abnormalities caused by disruption of hedgehog signaling. *Trends Genet* **28**: 364–373.

Andrey G, Montavon T, Mascrez B, Gonzalez F, Noordermeer D, Leleu M, Trono D, Spitz F, Duboule D. 2013. A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science* **340**: 1234167.

Andrey G, Mundlos S. 2017. The three-dimensional genome: regulating gene expression during pluripotency and development. *Development* **144**: 3646–3658.

Andrey G, Schöpflin R, Jerković I, Heinrich V, Ibrahim DM, Paliou C, Hochradel M, Timmermann B, Haas S, Vingron M, et al. 2017. Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding. *Genome Res* **27**: 223–233.

Andrey G, Spielmann M. 2017. CRISPR/Cas9 Genome Editing in Embryonic Stem Cells. In *Enhancer RNAs: Methods and protocols, Methods in Molecular Biology* (ed. U.A. Ørom), Vol. 1468 of, pp. 221–234, Springer New York, New York, NY

Arnold CD, Gerlach D, Stelzer C, Boryń ŁM, Rath M, Stark A. 2013. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**: 1074–7.).

Artus J, Hadjantonakis A-K. 2011. Generation of Chimeras by Aggregation of Embryonic Stem Cells with Diploid or Tetraploid Mouse Embryos. pp. 37–56, Humana Press

Bailey SD, Zhang X, Desai K, Aid M, Corradin O, Cowper-Sal Lari R, Akhtar-Zaidi B, Scacheri PC, Haibe-Kains B, Lupien M, et al. 2015. ZNF143 provides sequence specificity to secure chromatin interactions at gene promoters. *Nat Commun* **2**: 6186.

Banerji J, Rusconi S, Schaffner W. 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**: 299–308.

Baniahmad A, Steiner C, Köhne AC, Renkawitz R. 1990. Modular structure of a chicken lysozyme silencer: Involvement of an unusual thyroid hormone receptor binding site. *Cell* **61**: 505–514.

Barbieri M, Chotalia M, Fraser J, Lavitas L-M, Dostie J, Pombo A, Nicodemi M. 2012. Complexity of chromatin folding is captured by the strings and binders switch model. *Proc Natl Acad Sci* **109**: 16173–16178.

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P.

2007. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science (80-)* **315**: 1709–1712.

Barroso GV, Puzovic N, Dutheil JY. 2018. The Evolution of Gene-Specific Transcriptional Noise Is Driven by Selection at the Pathway Level. *Genetics* **208**: 173–189.

Bartman CR, Hsu SC, Hsiung CC-S, Raj A, Blobel GA. 2016. Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. *Mol Cell* **62**: 237–247.

Battulin N, Fishman VS, Mazur AM, Pomaznoy M, Khabarova AA, Afonnikov DA, Prokhortchouk EB, Serov OL. 2015. Comparison of the three-dimensional organization of sperm and fibroblast genomes using the Hi-C approach. *Genome Biol* **16**: 77.

Beagan JA, Duong MT, Titus KR, Zhou L, Cao Z, Ma J, Lachanski C V., Gillis DR, Phillips-Cremins JE. 2017. YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Res* **27**: 1139–1152.

Benabdallah NS, Gautier P, Hekimoglu-Balkan B, Lettice LA, Bhatia S, Bickmore WA. 2016. SBE6: a novel long-range enhancer involved in driving sonic hedgehog expression in neural progenitor cells. *Open Biol* **6**: 160197.

Benabdallah NS, Williamson I, Illingworth RS, Boyle S, Grimes GR, Therizols P, Bickmore W. 2017. PARP mediated chromatin unfolding is coupled to long-range enhancer activation. *bioRxiv*

Bénazet J-D, Bischofberger M, Tiecke E, Gonçalves A, Martin JF, Zuniga A, Naef F, Zeller R. 2009. A self-regulatory system of interlinked signaling feedback loops controls mouse limb patterning. *Science* **323**: 1050–3.

Bianco S, Lupiáñez DG, Chiariello AM, Annunziatella C, Kraft K, Schöpflin R, Wittler L, Andrey G, Vingron M, Pombo A, et al. 2018. Polymer physics predicts the effects of structural variants on chromatin architecture. *Nat Genet* **50**: 662–667.

Bibikova M, Beumer K, Trautman JK, Carroll D. 2003. Enhancing Gene Targeting with Designed Zinc Finger Nucleases. *Science (80-)* **300**: 764–764.

Bibikova M, Carroll D, Segal DJ, Trautman JK, Smith J, Kim YG, Chandrasegaran S. 2001. Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol* **21**: 289–97.

Bibikova M, Golic M, Golic K, Carroll D. 2002. Targeted Chromosomal Cleavage and Mutagenesis in Drosophila Using Zinc-Finger Nucleases. *Genetics* **161**: 1169–75.

Bickmore WA. 2013. The Spatial Organization of the Human Genome. www.annualreviews.org

Boffelli D, Nobrega MA, Rubin EM. 2004. Comparative genomics at the vertebrate extremes. *Nat Rev Genet* **5**: 456–465.

Bolotin A, Quinquis B, Sorokin A, Ehrlich SD. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**: 2551–2561.

Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, Xu X, Lv X, Hugnot J-PP, Tanay A, et al. 2017. Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**: 557-572.

Bothma JP, Garcia HG, Esposito E, Schlissel G, Gregor T, Levine M. 2014. Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living Drosophila embryos. *Proc Natl Acad Sci* **111**: 10598–10603.

Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE. 2008. High-Resolution Mapping and Characterization of Open Chromatin across the Genome. *Cell* **132**: 311–322.

Boyle S, Gilchrist S, Bridger JM, Mahy NL, Ellis JA, Bickmore WA. 2001. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum Mol Genet* **10**: 211–9.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**: 1213–8.

Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. 2015. ATAC-seq: A method for assaying chromatin

accessibility genome-wide. *Curr Protoc Mol Biol* **2015**: 21.29.1-21.29.9.

Busslinger GA, Stocsits R, van der Lelij P, Axelsson E, Tedeschi A, Galjart N, Peters J-M. 2017. Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nat Publ Gr* **544**.

C. elegans Sequencing Consortium TC elegans S. 1998. Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science* **282**: 2012–8.

Capellini TD, Di Giacomo G, Salsi V, Brendolan A, Ferretti E, Srivastava D, Zappavigna V, Selleri L. 2006. Pbx1/Pbx2 requirement for distal limb patterning is mediated by the hierarchical control of Hox gene spatial distribution and Shh expression. *Development* **133**: 2263–73.

Chathoth KT, Zabet NR. 2019. Chromatin architecture reorganisation during neuronal cell differentiation in Drosophila genome. *Genome Res* **44**.

Chen H, Johnson RL. 1999. Dorsoventral patterning of the vertebrate limb: a process governed by multiple events. *Cell Tissue Res* **296**: 67–73.

Chen H, Johnson RL. 2002. Interactions between dorsal-ventral patterning genes lmx1b, engrailed-1 and wnt-7a in the vertebrate limb. *Int J Dev Biol* **46**: 937–941.

Chen H, Levo M, Barinov L, Fujioka M, Jaynes JB, Gregor T. 2018. Dynamic interplay between enhancer–promoter topology and gene activity. *Nat Genet* **50**: 1296–1303.

Chiang C, Litingtung Y, Harris MP, Simandl BK, Li Y, Beachy PA, Fallon JF. 2001. Manifestation of the Limb Prepattern: Limb Development in the Absence of Sonic Hedgehog Function. *Dev Biol* **236**: 421–435.

Chiang C, Litingtung Y, Lee E, Young KE, Corden JL, Westphal H, Beachy P a. 1996. Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature* **383**: 407–413.

Chiariello AM, Annunziatella C, Bianco S, Esposito A, Nicodemi M. 2016. Polymer physics of chromosome large-scale 3D organisation. *Sci Rep* **6**: 29775.

Chiaromonte F, Weber RJ, Roskin KM, Diekhans M, Kent WJ, Haussler D. 2003. The share of human genomic DNA under selection estimated from human-mouse genomic alignments. *Cold Spring Harb Symp Quant Biol* **68**: 245–54.

Chiocchetti A, Tolosano E, Hirsch E, Silengo L, Altruda F. 1997. Green fluorescent protein as a reporter of gene expression in transgenic mice. *Biochim Biophys Acta - Gene Struct Expr* **1352**: 193–202.

Cho W-K, Spille J-H, Hecht M, Lee C, Li C, Grube V, Cisse II. 2018. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science (80-)* **361**: 412–415.

Christian M, Cermak T, Doyle EL, Schmidt C, Zhang F, Hummel A, Bogdanove AJ, Voytas DF. 2010. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**: 757–61.

Clarkson CT, Deeks EA, Samarista R, Zhurkin VB, Teif VB. 2019. CTCF binding strength modulates chromatin architecture through the changes of the local nucleosome repeat length. *bioRxiv*.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PPDPPDPPD, Wu X, Jiang W, Marraffini L a LA, et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science (80-)* **339**: 819–23.

Consortium MGS. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**: 520–562.

Cooper KL, Hu JK-H, ten Berge D, Fernández-Terán MA, Ros MÁ, Tabin CJ. 2011. Initiation of proximal-distal patterning in the vertebrate limb by signals and growth. *Science (80-)* **332**: 1083–1086.

Crawford GE, Holt IE, Whittle J, Webb BD, Tai D, Davis S, Margulies EH, Chen Y, Bernat JA, Ginsburg D, et al. 2005. Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). *Genome Res* **16**: 123–131.

Cremer T, Cremer C. 2001. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* **2**: 292–301.

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato M a, Frampton GM, Sharp P a, et al. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**: 21931–21936.

Crossley PH, Minowada G, MacArthur CA, Martin GR. 1996. Roles for FGF8 in the Induction, Initiation, and Maintenance of Chick Limb Development. *Cell* **84**: 127–136.

Cruz-Molina S, Respuela P, Tebartz C, Frommolt P, Cruz-molina S, Respuela P, Tebartz C, Kolovos P, Nikolic M, Fueyo R. 2017. PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation Article PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. *Stem Cell* 1–17.

Davies JOJ, Oudelaar AM, Higgs DR, Hughes JR. 2017. How best to identify chromosomal interactions : a comparison of approaches. *Nat Publ Gr* **14**.

de Wit E, Vos ESM, Holwerda SJB, Valdes-Quezada C, Verstegen MJAM, Teunissen H, Splinter E, Wijchers PJ, Krijger PHL, de Laat W. 2015. CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell* **60**: 676–684.

de Laat W, Duboule D. 2013. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502**: 499–506.

de Villiers J, Olson L, Tyndall C, Schaffner W. 1982. Transcriptional "enhancers" from SV40 and polyoma virus show a cell type preference. *Nucleic Acids Res* **10**: 7965–76.

Dekker J, Rippe K, Dekker M, Kleckner N. 2002. Capturing chromosome conformation. *Science* **295**: 1306–11.

Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**: 602–607.

Deng W, Lee J, Wang H, Miller J, Reik A, Gregory PD, Dean A, Blobel GA. 2012. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* **149**: 1233–44.

Deng W, Rupon JW, Krivega I, Breda L, Motta I, Jahn KS, Reik A, Gregory PD, Rivella S, Dean A, et al. 2014. Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell* **158**: 849–860.

Despang A, Schöpflin R, Franke M, Ali S, Jerković I, Paliou C, Chan W-L, Timmermann B, Wittler L, Vingron M, et al. 2019. Functional dissection of TADs reveals non-essential and instructive roles in regulating gene expression. *bioRxiv*.

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**: 376–80.

Dollé P, Dierich A, LeMeur M, Schimmang T, Schuhbaur B, Chambon P, Duboule D. 1993. Disruption of the Hoxd-13 gene induces localized heterochrony leading to mice with neotenic limbs. *Cell* **75**: 431–441.

Dorshorst B, Okimoto R, Ashwell C. 2010. Genomic Regions Associated with Dermal Hyperpigmentation, Polydactyly and Other Morphological Traits in the Silkie Chicken. *J Hered* **2010**: 339–350.

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. 2016. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**: 95–98.

Eckner R, Ewen ME, Newsome D, Gerdes M, DeCaprio JA, Lawrence JB, Livingston DM. 1994. Molecular cloning and functional analysis of the adenovirus E1A-associated 300-kD protein (p300) reveals a protein with properties of a transcriptional adaptor. *Genes Dev* **8**: 869–84.

Eijkelenboom A, Mokry M, de Wit E, Smits LM, Polderman PE, van Triest MH, van Boxtel R, Schulze A, de Laat W, Cuppen E, et al. 2013. Genome-wide analysis of FOXO3 mediated transcription regulation through RNA polymerase II profiling. *Mol Syst Biol* **9**: 638.

Epstein DJ, McMahon AP, Joyner AL. 1999. Regionalization of Sonic hedgehog transcription along the anteroposterior axis of the mouse central nervous system is regulated by Hnf3 dependent and independent mechanisms. *Development*.

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, et al. 2011. Mapping and analysis of chromatin state dynamics in nine human cell

types. *Nature* **473**: 43–9.

Ezkurdia I, Juan D, Rodriguez JM, Frankish A, Diekhans M, Harrow J, Vazquez J, Valencia A, Tress ML. 2014. Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Hum Mol Genet* **23**: 5866–78.

Flanagan PM, Kelleher RJ, Sayre MH, Tschochner H, Kornberg RD. 1991. A mediator required for activation of RNA polymerase II transcription in vitro. *Nature* **350**: 436–438.

Franke M, Ibrahim DM, Andrey G, Schwarzer W, Heinrich V, Schöpflin R, Kraft K, Kempfer R, Jerković I, Chan W-L, et al. 2016. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**.

Frankel N, Davis GK, Vargas D, Wang S, Payre F, Stern DL. 2010. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* **466**: 490–493.

Freire-Pritchett P, Schoenfelder S, Várnai C, Wingett SW, Cairns J, Collier AJ, García-Vílchez R, Furlan-Magaril M, Osborne CS, Fraser P, et al. 2017. Global reorganisation of cis-regulatory units upon lineage commitment of human embryonic stem cells. *Elife* **6**: 1–26.

Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31**: 822–826.

Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny Correspondence LA, Mirny LA. 2016. Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep* **15**: 2038–2049.

Fukaya T, Lim B, Levine M. 2016. Enhancer Control of Transcriptional Bursting. *Cell* **166**: 358–368.

Furlong EEM, Levine M. 2018. Developmental enhancers and chromosome topology. *Science* **361**: 1341–1345.

Galli A, Robay D, Osterwalder M, Bao X, Bénazet JD, Tariq M, Paro R, Mackem S, Zeller R. 2010. Distinct roles of Hand2 in initiating polarity and posterior Shh expression during the onset of mouse limb bud development. *PLoS Genet* **6**.

Gasiunas G, Barrangou R, Horvath P, Siksnys V. 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* **109**: E2579-86.

Gaulton KJ, Nammo T, Pasquali L, Simon JM, Giresi PG, Fogarty MP, Panhuis TM, Mieczkowski P, Secchi A, Bosco D, et al. 2010. A map of open chromatin in human pancreatic islets. *Nat Genet* **42**: 255–259.

Gehrke AR, Shubin NH. 2016. Cis-regulatory programs in the development and evolution of vertebrate paired appendages. *Semin Cell Dev Biol* **57**: 31–39.

Ghavi-Helm Y, Klein FA, Pakozdi T, Ciglar L, Noordermeer D, Huber W, Furlong EEM. 2014. Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**: 96–100.

Giorgetti L, Lajoie BR, Carter AC, Attia M, Zhan Y, Xu J, Chen CJ, Kaplan N, Chang HY, Heard E, et al. 2016. Structural organization of the inactive X chromosome in the mouse. *Nature* **535**: 575–579.

Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res* **17**: 877–885.

Guacci V, Koshland D, Strunnikov A. 1997. A Direct Link between Sister Chromatid Cohesion and Chromosome Condensation Revealed through the Analysis of MCD1 in S. cerevisiae. *Cell* **91**: 47–57.

Guo Y, Xu Q, Canzio D, Shou J, Li J, Gorkin DU, Jung I, Wu H, Zhai Y, Tang Y, et al. 2015. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* **162**: 900–910.

Haberle V, Stark A. 2018. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol* **19**: 1–17.

Hadjur S, Williams LM, Ryan NK, Cobb BS, Sexton T, Fraser P, Fisher AG, Merkenschlager M. 2009.

Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature* **460**: 410–413.

Haering CH, Farcas A-M, Arumugam P, Metson J, Nasmyth K. 2008. The cohesin ring concatenates sister DNA molecules. *Nature* **454**: 297–301.

Haft DH, Selengut J, Mongodin EF, Nelson KE. 2005. A Guild of 45 CRISPR-Associated (Cas) Protein Families and Multiple CRISPR/Cas Subtypes Exist in Prokaryotic Genomes. *PLoS Comput Biol* **1**: e60.

Hansen AS, Pustova I, Cattoglio C, Tjian R, Darzacq X. 2017. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife* **6**: e25776.

Harfe BD, Scherz PJ, Nissim S, Tian H, McMahon AP, Tabin CJ. 2004. Evidence for an Expansion-Based Temporal Shh Gradient in Specifying Vertebrate Digit Identities. *Cell* **118**: 517–528.

Heidari N, Phanstiel DH, He C, Grubert F, Jahanbani F, Kasowski M, Zhang MQ, Snyder MP. 2014. Genome-wide map of regulatory interactions in the human genome. *Genome Res* **24**: 1905–17.

Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, et al. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**: 108–12.

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, et al. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**: 311–318.

Heinz S, Texari L, Hayes MGB, Urbanowski M, Chang MW, Givarkes N, Rialdi A, White KM, Albrecht RA, Pache L, et al. 2018. Transcription Elongation Can Affect Genome 3D Structure. *Cell* **174**: 1522-1536.

Hesselberth JR, Chen X, Zhang Z, Sabo PJ, Sandstrom R, Reynolds AP, Thurman RE, Neph S, Kuehn MS, Noble WS, et al. 2009. Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nat Methods* **6**: 283–289.

Hill RE, Heaney SJH, Lettice LA. 2003. Sonic hedgehog: Restricted expression and limb dysmorphologies. *J Anat* **202**: 13–20.

Hnisz D, Shrinivas K, Young RA, Chakraborty AK, Sharp PA. 2017. A Phase Separation Model for Transcriptional Control. *Cell* **169**: 13–23.

Hobert O. 2010. Gene regulation: enhancers stepping out of the shadow. *Curr Biol* **20**: R697-9.

Hong J-W, Hendrix DA, Levine MS. 2008. Shadow Enhancers as a Source of Evolutionary Novelty. *Science (80-)* **321**: 1314.

Hou C, Li L, Qin ZS, Corces VG. 2012. Gene Density, Transcription, and Insulators Contribute to the Partition of the Drosophila Genome into Physical Domains. *Mol Cell* **48**: 471–484.

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, et al. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**: 827–32.

Hug CB, Grimaldi AG, Kruse K, Vaquerizas JM. 2017. Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription Correspondence Chromosome architecture gains complexity during development in concert with the onset of zygotic genome activation but without reliance on transcri. *Cell* **169**: 216–228.

Hughes JR, Roberts N, McGowan S, Hay D, Giannoulatou E, Lynch M, De Gobbi M, Taylor S, Gibbons R, Higgs DR. 2014. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat Genet* **46**: 205–212.

Ianakiev P, van Baren MJ, Daly MJ, Toledo SP, Cavalcanti MG, Neto JC, Silveira EL, Freire-Maia A, Heutink P, Kilpatrick MW, et al. 2001. Acheiropodia is caused by a genomic deletion in C7orf2, the human orthologue of the Lmbr1 gene. *Am J Hum Genet* **68**: 38–45.

Ibn-Salem J, Köhler S, Love MI, Chung H-R, Huang N, Hurles ME, Haendel M, Washington NL, Smedley D, Mungall CJ, et al. 2014. Deletions of chromosomal regulatory boundaries are associated with congenital disease. *Genome Biol* **15**: 423.

Ingham PW, McMahon AP. 2001. Hedgehog signaling in animal development: paradigms and principles. *Genes Dev* **15**: 3059–87.

Irimia M, Royo JL, Burguera D, Maeso I, Gómez-Skarmeta JL, Garcia-Fernandez J. 2012. Comparative genomics of the Hedgehog loci in chordates and the origins of Shh regulatory novelties. *Sci Rep* **2**: 433.

Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A. 1987. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol* **169**: 5429–33.

Jansen R, Embden JDA van, Gaastra W, Schouls LM. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **43**: 1565–1575.

Jeong Y, El-Jaick K, Roessler E, Muenke M, Epstein DJ. 2006. A functional screen for sonic hedgehog regulatory elements across a 1 Mb interval identifies long-range ventral forebrain enhancers. *Development* **133**: 761–772.

Jerković I, Ibrahim DM, Andrey G, Haas S, Hansen P, Janetzki C, Gonzá Lez Navarrete I, Robinson PN, Hecht J, Mundlos S, et al. 2017. Genome-Wide Binding of Posterior HOXA/D Transcription Factors Reveals Subgrouping and Association with CTCF ed. E.M. Mendenhall. *PLoS Genet* **13**: e1006567.

Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, Yen C-A, Schmitt AD, Espinoza CA, Ren B. 2013. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**: 290–294.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**: 816–21.

Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. 2013. RNA-programmed genome editing in human cells. *Elife* **2**: 471.

Jung YH, Sauria MEG, Lyu X, Cheema MS, Ausio J, Taylor J, Corces VG. 2017. Chromatin States in Mouse Sperm Correlate with Embryonic and Adult Regulatory Landscapes. *Cell Rep* **18**: 1366–1382.

Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, et al. 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**: 430–435.

Kalmar T, Lim C, Hayward P, Muñoz-Descalzo S, Nichols J, Garcia-Ojalvo J, Martinez Arias A. 2009. Regulated Fluctuations in Nanog Expression Mediate Cell Fate Decisions in Embryonic Stem Cells ed. M.A. Goodell. *PLoS Biol* **7**: e1000149.

Ke Y, Xu Y, Chen X, Xu X, Huang X, Correspondence JL. 2017. 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis. *Cell* **170**: 367–381.

Kelleher RJ, Flanagan PM, Kornberg RD. 1990. A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. *Cell* **61**: 1209–1215.

Kim S, Kim D, Cho SW, Kim J, Kim J-S. 2014a. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res* **24**: 1012–1019.

Kim T-H, Li F, Ferreiro-Neira I, Ho L-L, Luyten A, Nalapareddy K, Long H, Verzi M, Shivdasani RA. 2014b. Broadly permissive intestinal chromatin underlies lateral inhibition and cell plasticity. *Nature* **506**: 511–515.

Klopocki E, Mundlos S. 2011. Copy-number variations, noncoding sequences, and human phenotypes. *Annu Rev Genomics Hum Genet* **12**: 53–72.

Kmita M, Tarchini B, Zàkàny J, Logan M, Tabin CJ, Duboule D. 2005. Early developmental arrest of mammalian limbs lacking HoxA/HoxD gene function. *Nature* **435**: 1113–1116.

Knight PA, Ruiz D. 2013. A fast algorithm for matrix balancing. *IMA J Numer Anal* **33**: 1029–1047.

Kondo T, Duboule D. 1999. Breaking colinearity in the mouse HoxD complex. *Cell* **97**: 407–17.

Kothary R, Clapoff S, Darling S, Perry MD, Moran LA, Rossant J. 1989. Inducible expression of an hsp68-lacZ hybrid gene in transgenic mice. *Development* **105**: 707–714.

Kraft K, Geuer S, Will AJ, Chan W, Paliou C, Borschiwer M, Harabula I, Wittler L, Franke M, Ibrahim DM, et al. 2015. Deletions, inversions, duplications: Engineering of structural variants using CRISPR/Cas in mice. *Cell Rep* **10**: 833–839.

Kragesteen BK, Spielmann M, Paliou C, Heinrich V, Schöpflin R, Esposito A, Annunziatella C, Bianco S, Chiariello AM, Jerković I, et al. 2018. Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. *Nat Genet* **50**: 1463–1473.

Krebs O, Schreiner CM, Scott WJ, Bell SM, Robbins DJ, Goetz J a, Alt H, Hawes N, Wolf E, Favor J. 2003. Replicated anterior zeugopod (raz): a polydactylous mouse mutant with lowered Shh signaling in the limb bud. *Development* **130**: 6037–6047.

Kvon EZ. 2015. Using transgenic reporter assays to functionally characterize enhancers in animals. *Genomics* **106**: 185–192.

Kvon EZ, Kamneva OK, Melo US, Barozzi I, Osterwalder M, Mannion BJ, Tissi??res V, Pickle CS, Plajzer-Frick I, Lee EA, et al. 2016. Progressive Loss of Function in a Limb Enhancer during Snake Evolution. *Cell* **167**: 633-642.e11.

Lander E. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359.

Larsson AJM, Johnsson P, Hagemann-Jensen M, Hartmanis L, Faridani OR, Reinius B, Segerstolpe Å, Rivera CM, Ren B, Sandberg R. 2019. Genomic encoding of transcriptional burst kinetics. *Nature* **565**.

Leal F, Cohn MJ. 2016. Loss and Re-emergence of Legs in Snakes by Modular Evolution of Sonic hedgehog and HOXD Enhancers. *Curr Biol* **26**: 2966–2973.

Lee T, Johnston S, Young R. 2006. Chromatin immunoprecipitation from C. elegans embryos. *Nat Protoc* **1**: 729–748.

Lee W, Haslinger A, Karin M, Tjian R. 1987. Activation of transcription by two factors that bind promoter and enhancer sequences of the human metallothionein gene and SV40. *Nature* **325**: 368–372.

Lettice L a, Horikoshi T, Heaney SJH, van Baren MJ, van der Linde HC, Breedveld GJ, Joosse M, Akarsu N, Oostra B a, Endo N, et al. 2002. Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. *Proc Natl Acad Sci U S A* **99**: 7548–53.

Lettice L a, Williamson I, Devenney PS, Kilanowski F, Dorin J, Hill RE. 2014. Development of five digits is controlled by a bipartite long-range cis-regulator. *Development* **141**: 1715–25.

Lettice LA, Daniels S, Sweeney E, Venkataraman S, Devenney PS, Gautier P, Morrison H, Fantes J, Hill RE, Fitzpatrick DR. 2011. Enhancer-adoption as a mechanism of human developmental disease. *Hum Mutat* **32**: 1492–1499.

Lettice LA, Devenney P, De Angelis C, Correspondence REH, Hill RE. 2017. The Conserved Sonic Hedgehog Limb Enhancer Consists of Discrete Functional Elements that Regulate Precise Spatial Expression. *CellReports* **20**: 1396–1408.

Lettice LA, Heaney SJH, Purdie LA, Li L, de Beer P, Oostra BA, Goode D, Elgar G, Hill RE, de Graaff E. 2003. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum Mol Genet* **12**: 1725–1735.

Lettice LA, Hill AE, Devenney PS, Hill RE. 2008. Point mutations in a distant sonic hedgehog cis-regulator generate a variable regulatory output responsible for preaxial polydactyly. *Hum Mol Genet* **17**: 978–985.

Lettice LA, Williamson I, Wiltshire JH, Peluso S, Devenney PS, Hill AE, Essafi A, Hagman J, Mort R, Grimes G, et al. 2012. Opposing functions of the ETS factor family define Shh spatial expression in limb buds and underlie polydactyly. *Dev Cell* **22**: 459–67.

Levine M, Cattoglio C, Tjian R. 2014. Looping back to leap forward: transcription enters a new era. *Cell* **157**: 13–25.

Lewandoski M, Sun X, Martin GR. 2000. Fgf8 signalling from the AER is essential for normal limb development. *Nat Genet* **26**: 460–463.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Lichter P, Cremer T, Borden J, Manuelidis L, Ward DC. 1988. Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using

recombinant DNA libraries. *Hum Genet* **80**: 224–34.
Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**: 289–93.
Lin S, Staahl BT, Alla RK, Doudna JA. 2014. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3**.
Ling JQ, Li T, Hu JF, Vu TH, Chen HL, Qiu XW, Cherry AM, Hoffman AR. 2006. CTCF Mediates Interchromosomal Colocalization Between Igf2/H19 and Wsb1/Nf1. *Science (80-)* **312**: 269–272.
Livak KJ, Schmittgen TD. 2001. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods* **25**: 402–408.
Lobanenkov V V, Nicolas RH, Adler V V, Paterson H, Klenova EM, Polotskaja A V, Goodwin GH. 1990. A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. *Oncogene* **5**: 1743–53.
Loomis CA, Michaud J, Joyner AL, Wurst W, Harris E, Hanks M. 2003. The mouse Engrailed-1 gene and ventral limb patterning. *Nature* **382**: 360–363.
Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, et al. 2015. Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* **161**: 1–14.
Lupiáñez DG, Spielmann M, Mundlos S. 2016. Breaking TADs: How Alterations of Chromatin Domains Result in Disease. *Trends Genet* **32**: 225–237.
Makarova KS, Grishin N V, Shabalina SA, Wolf YI, Koonin E V. 2006. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct* **1**: 7.
Makarova KS, Haft DH, Barrangou R, Brouns SJJ, Charpentier E, Horvath P, Moineau S, Mojica FJM, Wolf YI, Yakunin AF, et al. 2011. Evolution and classification of the CRISPR–Cas systems. *Nat Rev Microbiol* **9**: 467–477.
Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. 2013. RNA-guided human genome engineering via Cas9. *Science* **339**: 823–6.
Mariani F V., Ahn CP, Martin GR. 2008. Genetic evidence that FGFs have an instructive role in limb proximal–distal patterning. *Nature* **453**: 401–405.
Marinić M, Aktas T, Ruf S, Spitz F. 2013. An Integrated Holo-Enhancer Unit Defines Tissue and Gene Specificity of the Fgf8 Regulatory Landscape. *Dev Cell* **24**: 530–542.
Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**: 10.
Matsumaru D, Haraguchi R, Miyagawa S, Motoyama J, Nakagata N, Meijlink F, Yamada G. 2011. Genetic analysis of hedgehog signaling in ventral body wall development and the onset of omphalocele formation. *PLoS One* **6**: 9–12.
Medina-Rivera A, Santiago-Algarra D, Puthier D, Spicuglia S. 2018. Widespread Enhancer Activity from Core Promoters.
Mercader N, Leonardo E, Piedra ME, Martínez-A C, Ros MA, Torres M. 2000. Opposing RA and FGF signals control proximodistal vertebrate limb development through regulation of Meis genes. *Development* **127**: 3961–70.
Michaelis C, Ciosk R, Nasmyth K. 1997. Cohesins: Chromosomal Proteins that Prevent Premature Separation of Sister Chromatids. *Cell* **91**: 35–45.
Michos O, Panman L, Vintersten K, Beier K, Zeller R, Zuniga A. 2004. Gremlin-mediated BMP antagonism induces the epithelial-mesenchymal feedback signaling controlling metanephric kidney and limb organogenesis. *Development* **131**: 3401–10.
Mojica FJM, Díez-Villaseñor C, García-Martínez J, Soria E. 2005. Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *J Mol Evol* **60**: 174–182.

Mojica FJM, Diez-Villasenor C, Soria E, Juez G. 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol Microbiol* **36**: 244–246.

Montavon T, Soshnikova N, Mascrez B, Joye E, Thevenet L, Splinter E, de Laat W, Spitz F, Duboule D. 2011. A regulatory archipelago controls Hox genes transcription in digits. *Cell* **147**: 1132–45.

Muerdter F, Boryń ŁM, Woodfin AR, Neumayr C, Rath M, Zabidi MA, Pagani M, Haberle V, Kazmar T, Catarino RR, et al. 2018. Resolving systematic errors in widely used enhancer activity assays in human cells. *Nat Methods* **15**: 141–149.

Mundlos S, Horn D. 2014. Development of the Limbs. In *Limb Malformations*, pp. 2–9, Springer Berlin Heidelberg, Berlin, Heidelberg.

Narendra V, Bulajić M, Dekker J, Mazzoni EO, Reinberg D. 2016. CTCF-mediated topological boundaries during development foster appropriate gene regulation. *Genes Dev* **30**: 2657–2662.

Narendra V, Rocha PP, An D, Raviram R, Skok JA, Mazzoni EO, Reinberg D. 2015. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* **347**: 1017–21.

Nasmyth K. 2001. Disseminating the Genome: Joining, Resolving, and Separating Sister Chromatids During Mitosis and Meiosis. *Annu Rev Genet* **35**: 673–745.

Niederreither K, Vermot J, Schuhbaur B, Chambon P, Dollé P. 2002. Embryonic retinoic acid synthesis is required for forelimb growth and anteroposterior patterning in the mouse. *Development* **124**: 4225–4234.

Nir G, Farabella I, Pérez Estrada C, Ebeling CG, Beliveau BJ, Sasaki HM, Lee SD, Nguyen SC, McCole RB, Chattoraj S, et al. 2018. Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling ed. G.P. Copenhaver. *PLOS Genet* **14**: e1007872.

Nissim S, Hasso SM, Fallon JF, Tabin CJ. 2006. Regulation of Gremlin expression in the posterior limb bud. *Dev Biol* **299**: 12–21.

Nobrega MA, Ovcharenko I, Afzal V, Rubin EM. 2003. Scanning Human Gene Deserts for Long-Range Enhancers. *Science (80-)* **302**: 413.

Nora EP, Goloborodko A, Valton A-L, Dekker J, Mirny LA, Bruneau Correspondence BG. 2017. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**: 930–944.

Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, et al. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**: 381–5.

Nüsslein-Volhard C, Wieschaus E. 1980. Mutations affecting segment number and polarity in Drosophila. *Nature* **287**: 795–801.

O'Kane CJ, Gehring WJ. 1987. Detection in situ of genomic regulatory elements in Drosophila. *Proc Natl Acad Sci U S A* **84**: 9123–7.

Ohuchi H, Nakagawa T, Yamamoto A, Araga A, Ohata T, Ishimaru Y, Yoshioka H, Kuwana T, Nohno T, Yamasaki M, et al. 1997. The mesenchymal factor, FGF10, initiates and maintains the outgrowth of the chick limb bud through interaction with FGF8, an apical ectodermal factor. *Development* **124**: 2235–44.

Ong C-T, Corces VG. 2014. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet* **15**: 234–246.

Osterwalder M, Barozzi I, Tissières V, Fukuda-Yuzawa Y, Mannion BJ, Afzal SY, Lee EA, Zhu Y, Plajzer-Frick I, Pickle CS, et al. 2018. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nat Publ Gr* **554**.

Osterwalder M, Speziale D, Shoukry M, Mohan R, Ivanek R, Kohler M, Beisel C, Wen X, Scales SJ, Christoffels VM, et al. 2014. HAND2 targets define a network of transcriptional regulators that compartmentalize the early limb bud mesenchyme. *Dev Cell* **31**: 345–357.

Paliou C, Guckelberger P, Schöpflin R, Heinrich V, Esposito A, Chiariello AM, Bianco S, Annunziatella C, Helmuth J, Haas S, et al. 2019. Preformed chromatin topology assists transcriptional

robustness of Shh during limb development. *Proc Natl Acad Sci U S A* 201900672.

Parelho V, Hadjur S, Spivakov M, Leleu M, Sauer S, Gregson HC, Jarmuz A, Canzonetta C, Webster Z, Nesterova T, et al. 2008. Cohesins Functionally Associate with CTCF on Mammalian Chromosome Arms. *Cell* **132**: 422–433.

Parr BA, McMahon AP. 1995. Dorsalizing signal Wnt-7a required for normal polarity of D–V and A–P axes of mouse limb. *Nature* **374**: 350–353.

Pasquali L, Gaulton KJ, Rodríguez-Seguí SA, Mularoni L, Miguel-Escalada I, Akerman İ, Tena JJ, Morán I, Gómez-Marín C, van de Bunt M, et al. 2014. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat Genet* **46**: 136–143.

Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR. 2013. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* **31**: 839–843.

Peng Y, Clark KJ, Campbell JM, Panetta MR, Guo Y, Ekker SC. 2014. Making designer mutants in model organisms. *Development* **141**: 4042–4054.

Pennacchio L a, Bickmore W, Dean A, Nobrega MA, Bejerano G. 2013. Enhancers: five essential questions. *Nat Rev Genet* **14**: 288–95.

Petit F, Sears KE, Ahituv N. 2017. Limb development: a paradigm of gene regulation. *Nat Rev Genet*.

Phillips-Cremins JE, Sauria MEG, Sanyal A, Gerasimova TI, Lajoie BR, Bell JSK, Ong C-T, Hookway TA, Guo C, Sun Y, et al. 2013. Architectural Protein Subclasses Shape 3D Organization of Genomes during Lineage Commitment. *Cell* **153**: 1281–1295.

Pinkel D, Landegent J, Collins C, Fuscoe J, Segraves R, Lucas J, Gray J. 1988. Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. *Proc Natl Acad Sci* **85**: 9138–9142.

Pourcel C, Salvignol G, Vergnaud G. 2005. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151**: 653–663.

Pradeepa MM, Grimes GR, Kumar Y, Olley G, Taylor GCA, Schneider R, Bickmore WA. 2016. Histone H3 globular domain acetylation identifies a new class of enhancers. *Nat Genet* **48**: 681–686.

Racko D, Benedetti F, Dorier J, Stasiak A. 2018. Transcription-induced supercoiling as the driving force of chromatin loop extrusion during formation of TADs in interphase chromosomes. *Nucleic Acids Res* **46**: 1648–1660.

Rada-Iglesias A, Bajpai R, Swigut T, Brugmann S a, Flynn R a, Wysocka J. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**: 279–83.

Rao SSP, Huang S-C, Glenn B, Hilaire S, Casellas R, Lander ES, Lieberman E, Correspondence A. 2017. Cohesin Loss Eliminates All Loop Domains. *Cell* **171**: 305–320.

Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**: 1665–1680.

Ren G, Jin W, Cui K, Rodrigez J, Hu G, Zhang Z, Larson DR, Zhao K. 2017. CTCF-Mediated Enhancer-Promoter Interaction Is a Critical Regulator of Cell-to-Cell Variation of Gene Expression. *Mol Cell* **67**: 1049-1058.

Reyon D, Tsai SQ, Khayter C, Foden JA, Sander JD, Joung JK. 2012. FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol* **30**: 460–465.

Riddle RD, Ensini M, Nelson C, Tsuchida T, Jessell TM, Tabin C. 1995. Induction of the LIM homeobox gene Lmx1 by WNT7a establishes dorsoventral pattern in the vertebrate limb. *Cell* **83**: 631–40.

Riddle RD, Johnson RL, Laufer E, Tabin C. 1993. Sonic hedgehog mediates the polarizing activity of the ZPA. *Cell* **75**: 1401–1416.

Rosello-Diez A, Ros MA, Torres M. 2011. Diffusible Signals, Not Autonomous Mechanisms, Determine the Main Proximodistal Limb Subdivision. *Science (80-)* **332**: 1086–1088.

Rouet P, Smih F, Jasin M. 1994. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol* **14**: 8096–106.

Rowley MJ, Nichols MH, Lyu X, Ando-Kuri M, Rivera ISM, Hermetz K, Wang P, Ruan Y, Corces VG. 2017. Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol Cell* **67**: 837–852.

Rubin AJ, Barajas BC, Furlan-Magaril M, Lopez-Pajares V, Mumbach MR, Howard I, Kim DS, Boxer LD, Cairns J, Spivakov M, et al. 2017a. Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. *Nat Genet* **49**: 1522–1528.

Rubin AJ, Barajas BC, Furlan-Magaril M, Lopez-Pajares V, Mumbach MR, Howard I, Kim DS, Boxer LD, Cairns J, Spivakov M, et al. 2017b. Lineage-specific dynamic and pre-established enhancer–promoter contacts cooperate in terminal differentiation. *Nat Genet* **49**: 1522–1528.

Rubio ED, Reiss DJ, Welcsh PL, Disteche CM, Filippova GN, Baliga NS, Aebersold R, Ranish JA, Krumm A. 2008. CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci U S A* **105**: 8309–14.

Ruf S, Symmons O, Uslu VV, Dolle D, Hot C, Ettwiller L, Spitz F. 2011. Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. *Nat Genet* **43**: 379–86.

Sagai T, Amano T, Tamura M, Mizushina Y, Sumiyama K, Shiroishi T. 2009. A cluster of three long-range enhancers directs regional Shh expression in the epithelial linings. *Development* **136**: 1665–1674.

Sagai T, Hosoya M, Mizushina Y, Tamura M, Shiroishi T. 2005. Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. *Development* **132**: 797–803.

Sambrook J, Russell DW. 2001. Molecular Cloning - Sambrook & Russel - Vol. 1, 2, 3. *Cold Spring Harb Lab Press* **3th Edition**.

Samstein RM, Arvey A, Josefowicz SZ, Peng X, Reynolds A, Sandstrom R, Neph S, Sabo P, Kim JM, Liao W, et al. 2012. Foxp3 Exploits a Pre-Existent Enhancer Landscape for Regulatory T Cell Lineage Specification. *Cell* **151**: 153–166.

Sanborn AL, Rao SSP, Huang S-C, Durand NC, Huntley MH, Jewett AI, Bochkov ID, Chinnappan D, Cutkosky A, Li J, et al. 2015. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci* **112**: E6456–E6465.

Sander JD, Joung JK. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol*.

Sanyal A, Lajoie BR, Jain G, Dekker J. 2012. The long-range interaction landscape of gene promoters. *Nature* **489**: 109–113.

Schwalie PC, Ward MC, Cain CE, Faure AJ, Gilad Y, Odom DT, Flicek P. 2013. Co-binding by YY1 identifies the transcriptionally active, highly conserved set of CTCF-bound regions in primate genomes. *Genome Biol* **14**: R148.

Schwarzer W, Abdennur N, Goloborodko A, Pekowska A, Fudenberg G, Loe-mie Y, Fonseca nuno A, Huber W, Haering christian H, Mirny L, et al. 2017. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**.

Sears KE, Maier JA, Rivas-Astroza M, Poe R, Zhong S, Kosog K, Marcot JD, Behringer RR, Cretekos CJ, Rasweiler JJ, et al. 2015. The Relationship between Gene Network Structure and Expression Variation among Individuals and Species ed. A. Kopp. *PLOS Genet* **11**: e1005398.

Seitan VC, Faure AJ, Zhan Y, McCord RP, Lajoie BR, Ing-Simmons E, Lenhard B, Giorgetti L, Heard E, Fisher AG, et al. 2013. Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. *Genome Res* **23**: 2066–77.

Sexton T, Schober H, Fraser P, Gasser SM. 2007. Gene regulation through nuclear organization. *Nat Struct Mol Biol* **14**: 1049–1055.

Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G. 2012. Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell* **148**: 458–472.

Shah SA, Erdmann S, Mojica FJM, Garrett RA. 2013. Protospacer recognition motifs. *RNA Biol* **10**:

891–899.
Sharpe J, Lettice L, Hecksher-Sørensen J, Fox M, Hill R, Krumlauf R. 1999. Identification of Sonic hedgehog as a candidate gene responsible for the polydactylous mouse mutant Sasquatch. *Curr Biol* **9**: 97-S1.
Simonis M, Klous P, Homminga I, Galjaard R-J, Rijkers E-J, Grosveld F, Meijerink JPP, de Laat W. 2009. High-resolution identification of balanced and complex chromosomal rearrangements by 4C technology. *Nat Methods* **6**: 837–842.
Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, De Wit E, Van Steensel B, De Laat W. 2006. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* **38**: 1348–1354.
Sofueva S, Yaffe E, Chan W-C, Georgopoulou D, Vietri Rudan M, Mira-Bontenbal H, Pollard SM, Schroth GP, Tanay A, Hadjur S. 2013. Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J* **32**: 3119–3129.
Spilianakis CG, Flavell RA. 2004. Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nat Immunol* **5**: 1017–1027.
Splinter E, Heath H, Kooren J, Palstra R-J, Klous P, Grosveld F, Galjart N, de Laat W. 2006. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev* **20**: 2349–54.
Stadhouders R, Kolovos P, Brouwer R, Zuin J, van den Heuvel A, Kockx C, Palstra R-J, Wendt KS, Grosveld F, van Ijcken W, et al. 2013. Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nat Protoc* **8**: 509–524.
Stevens TJ, Lando D, Basu S, Atkinson LP, Cao Y, Lee SF, Leeb M, Wohlfahrt KJ, Boucher W, O'Shaughnessy-Kirwan A, et al. 2017. 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature*.
Symmons O, Pan L, Remeseiro S, Aktas T, Klein F, Huber W, Spitz F, Amano T, Sagai T, Tanabe H, et al. 2016. The Shh Topological Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic Distances. *Dev Cell* **0**: 47–57.
Symmons O, Uslu VV, Tsujimura T, Ruf S, Nassari S, Schwarzer W, Ettwiller L, Spitz F. 2014. Functional and topological characteristics of mammalian regulatory domains. *Genome Res* **24**: 390–400.
Szabo Q, Bantignies F, Cavalli G. 2019. Principles of genome folding into topologically associating domains. *Sci Adv* **5**: eaaw1668.
Tanabe H, Habermann FA, Solovei I, Cremer M, Cremer T. 2002. Non-random radial arrangements of interphase chromosome territories: evolutionary considerations and functional implications. *Mutat Res Mol Mech Mutagen* **504**: 37–45.
Tarchini B, Duboule D. 2006. Control of Hoxd Genes' Collinearity during Early Limb Development. *Dev Cell* **10**: 93–103.
Tarchini B, Duboule D, Kmita M. 2006. Regulatory constraints in the evolution of the tetrapod limb anterior–posterior polarity. *Nature* **443**: 985–988.
te Welscher P, Zuniga A, Kuijper S, Drenth T, Goedemans HJ, Meijlink F, Zeller R. 2002. Progression of Vertebrate Limb Development Through SHH-Mediated Counteraction of GLI3. *Science (80-)* **298**: 827–830.
Tickle C. 1981. The number of polarizing region cells required to specify additional digits in the developing chick wing. *Nature* **289**: 295–8.
Tickle C, Towers M. 2017. Sonic Hedgehog Signaling in Limb Development. *Front cell Dev Biol* **5**: 14.
Tolhuis B, Palstra RJ, Splinter E, Grosveld F, De Laat W. 2002. Looping and interaction between hypersensitive sites in the active β-globin locus. *Mol Cell* **10**: 1453–1465.
Tsukiji N, Amano T, Shiroishi T. 2014. A novel regulatory element for Shh expression in the lung and gut of mouse embryos. *Mech Dev* **131**: 127–136.
Uhlmann F, Nasmyth K. 1998. Cohesion between sister chromatids must be established during DNA

replication. *Curr Biol* **8**: 1095–101.

Ulianov S V, Khrameeva EE, Gavrilov AA, Flyamer IM, Kos P, Mikhaleva EA, Penin AA, Logacheva MD, Imakaev M V, Chertovich A, et al. 2016. Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res* **26**: 70–84.

Urban EA, Johnston RJ. 2018. Buffering and Amplifying Transcriptional Noise During Cell Fate Specification. *Front Genet* **9**: 591.

van de Werken HJG, de Vree PJP, Splinter E, Holwerda SJB, Klous P, de Wit E, de Laat W. 2012. 4C Technology: Protocols and Data Analysis. *Methods Enzymol* **513**: 89–112.

van Steensel B, Furlong EEM. 2019. The role of transcription in shaping the spatial organization of the genome. *Nat Rev Mol Cell Biol* 1.

Vietri Rudan M, Barrington C, Henderson S, Ernst C, Odom DT, Tanay A, Hadjur S. 2015. Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture. *Cell Rep* **10**: 1297–1309.

Visel A, Blow MJ, Li Z, Zhang T, Akiyama J a, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, et al. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**: 854–8.

Vogel A, Rodriguez C, Warnken W, Belmonte JCI. 1995. Dorsal cell fate specified by chick Lmxl during vertebrate limb development. *Nature* **378**: 716–720.

Wang B, Fallon JF, Beachy PA. 2000. Hedgehog-Regulated Processing of Gli3 Produces an Anterior/Posterior Repressor Gradient in the Developing Vertebrate Limb. *Cell* **100**: 423–434.

Wang S, Su J-H, Beliveau BJ, Bintu B, Moffitt JR, Wu C, Zhuang X. 2016. Spatial organization of chromatin domains and compartments in single chromosomes. *Science* **353**: 598–602.

Weintraub AS, Li CH, Zamudio A V, Bradner JE, Gray NS, Young Correspondence RA, Sigova AA, Hannett NM, Day DS, Abraham BJ, et al. 2017. YY1 Is a Structural Regulator of Enhancer-Promoter Loops YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* **172**.

Weischenfeldt J, Dubash T, Drainas AP, Mardin BR, Chen Y, Stütz AM, Waszak SM, Bosco G, Halvorsen AR, Raeder B, et al. 2017. Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. *Nat Genet* **49**: 65–74.

Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara K, Mishiro T, et al. 2008. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**: 796–801.

Wieczorek D, Pawlik B, Li Y, Akarsu NA, Caliebe A, May KJW, Schweiger B, Vargas FR, Balci S, Gillessen-Kaesbach G, et al. 2010. A specific mutation in the distant sonic hedgehog (SHH) cis-regulator (ZRS) causes Werner Mesomelic Syndrome (WMS) while complete ZRS duplications underlie Haas type polysyndactyly and preaxial polydactyly (PPD) with or without triphalangeal thumb. *Hum Mutat* **31**: 81–89.

Will AJ, Cova G, Osterwalder M, Chan W-L, Wittler L, Brieske N, Heinrich V, de Villartay J-P, Vingron M, Klopocki E, et al. 2017. Composition and dosage of a multipartite enhancer cluster control developmental expression of Ihh (Indian hedgehog). *Nat Genet* **49**: 1539–1545.

Williamson I, Kane L, Devenney PS, Anderson E, Kilanowski F, Hill RE, Bickmore WA, Lettice LA, Road C, Technology PA-T, et al. 2019. Developmentally regulated Shh expression is robust to TAD perturbations. *bioRxiv*.

Williamson I, Lettice LA, Hill RE, Bickmore WA. 2016. Shh and ZRS enhancer colocalisation is specific to the zone of polarising activity. *Development* **143**: 2994–3001.

Wingett S, Ewels P, Furlan-Magaril M, Nagano T, Schoenfelder S, Fraser P, Andrews S. 2015. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Research* **4**: 1310.

Wutz G, Várnai C, Nagasaka K, Cisneros DA, Stocsits RR, Tang W, Schoenfelder S, Jessberger G, Muhar M, Hossain MJ, et al. 2017. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J* **36**: 3573–3599.

Xu B, Wellik DM. 2011. Axial Hox9 activity establishes the posterior field in the developing forelimb. *Proc Natl Acad Sci U S A* **108**: 4888–91.

Yao TP, Oh SP, Fuchs M, Zhou ND, Ch'ng LE, Newsome D, Bronson RT, Li E, Livingston DM, Eckner R. 1998. Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. *Cell* **93**: 361–72.

Yao Y, Minor PJ, Zhao Y-T, Jeong Y, Pani AM, King AN, Symmons O, Gan L, Cardoso W V, Spitz F, et al. 2016. Cis-regulatory architecture of a brain signaling center predates the origin of chordates. *Nat Genet* **48**: 575–580.

Yuan G-C, Liu Y-J, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ. 2005. Genome-Scale Identification of Nucleosome Positions in S. cerevisiae. *Science (80-)* **309**: 626–630.

Zabidi M a., Arnold CD, Schernhuber K, Pagani M, Rath M, Frank O, Stark A. 2014. Enhancer––core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518**: 556–559.

Zabidi MA, Stark A. 2016. Regulatory Enhancer-Core-Promoter Communication via Transcription Factors and Cofactors. *Trends Genet* **32**: 801–814.

Zentner GE, Tesar PJ, Scacheri PC. 2011. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res* **21**: 1273–83.

Zhang F, Cong L, Lodato S, Kosuri S, Church GM, Arlotta P. 2011. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol* **29**: 149–53.

Zhao Z, Tavoosidana G, Sjölinder M, Göndör A, Mariano P, Wang S, Kanduri C, Lezcano M, Singh Sandhu K, Singh U, et al. 2006. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet* **38**: 1341–1347.

Zhu J, Mackem S. 2017. John Saunders' ZPA, Sonic hedgehog and digit identity – How does it really all work? *Dev Biol* **429**: 391–400.

Zlatanova J, Caiafa P. 2009. CTCF and its protein partners: divide and rule? *J Cell Sci* **122**: 1275–84.

Zuin J, Dixon JR, Reijden MIJA van der, Ye Z, Kolovos P, Brouwer RWW, Corput MPC van de, Werken HJG van de, Knoch TA, IJcken WFJ van, et al. 2014. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci* **111**: 996–1001.

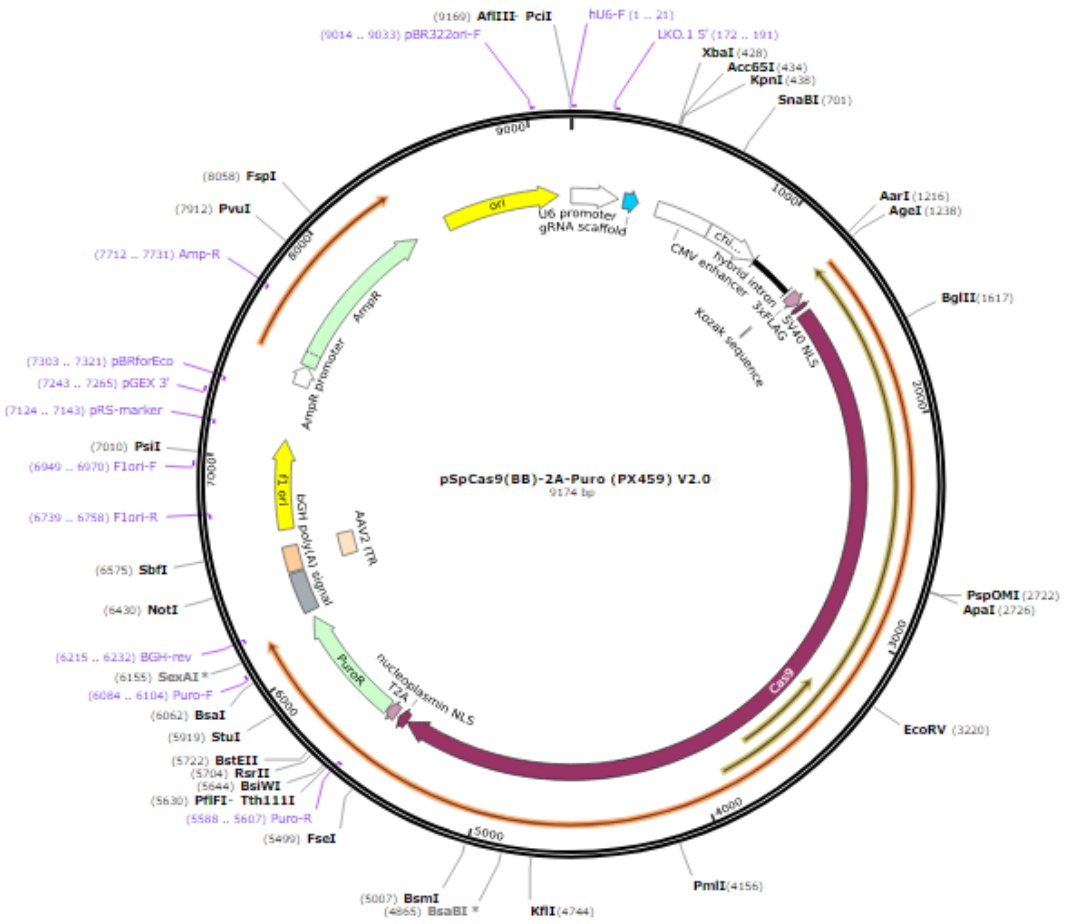
Zúñiga A, Haramis A-PG, McMahon AP, Zeller R. 1999. Signal relay by BMP antagonism controls the SHH/FGF4 feedback loop in vertebrate limb buds. *Nature* **401**: 598–602.

Zuniga A, Zeller R. 1999. Gli3 (Xt) and formin (ld) participate in the positioning of the polarising region and control of posterior limb-bud identity. *Development* **126**.

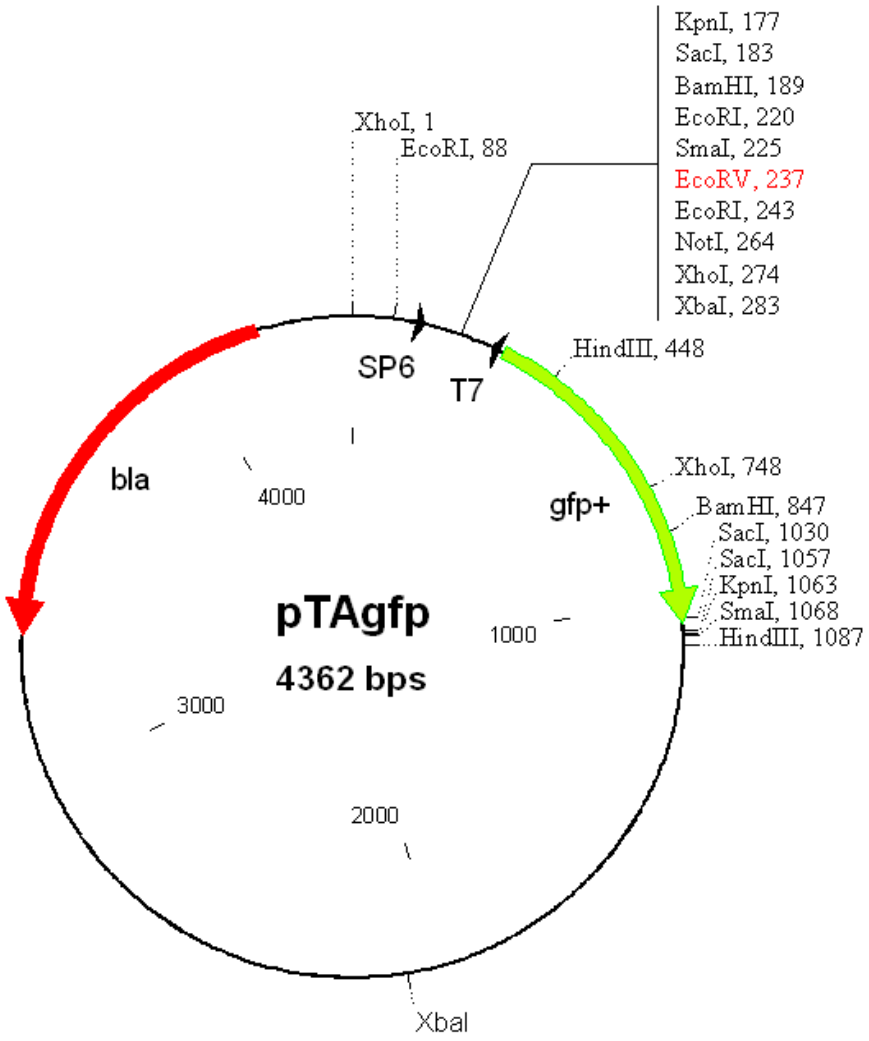
9 APPENDIX

9.1 Vector maps

pSpCas9 (BB)-2A-Puro (PX459)



pTA-GFP


KpnI, 177
SacI, 183
BamHI, 189
EcoRI, 220
SmaI, 225
EcoRV, 237
EcoRI, 243
NotI, 264
XhoI, 274
XbaI, 283
XhoI, 1
EcoRI, 88
HindIII, 448
SP6
T7
XhoI, 748
BamHI, 847
SacI, 1030
SacI, 1057
KpnI, 1063
SmaI, 1068
HindIII, 1087
bla
gfp+
4000
1000
2000
3000
pTAgfp
4362 bps
XbaI

9.2 Supplementary Tables

Table 9. 1. Single gRNA sequences

NAME	SEQUENCE (5'--> 3')
Lmbr1_3'_F	GTTTCCTGGACAACCGCGTTC
Lmbr1_5'_F:	GAGTGTAGACAAGTCTTTCGT
gCTCF_Lmbr1_i4_F	tattttgtgataccagcagg
gCTCF_Lmbr1_i5_F	ggagtcctctagtggccaac
gCTCF_ZRS_F	GAAACCAGTGCTCCCTAGTG
ZRS_Reg_CenF	GAAACCAGTGCTCCCTAGTG
ZRS_Reg_TelF	CTGAGACAAATTAGCCACTG
CTCF_Lmbr1_i9_cenF	AGGGCGTCAGGAAATTCCAC
CTCF_Lmbr1_i9_telF	TGAACTGCCAATCACCTGGG

Table 9. 2. Primers used for genotyping clones

NAME	SEQUENCE (5'--> 3')	Tested Breakpoint
Typ32	CCTGACCTGCTAGAAACAGC	Lmbr1_prom_CenF
Typ33	GGGTATAGGGTGATGGGTGG	Lmbr1_prom_CenR
Typ34	GGCTCATCAGTCAGCAACAA	Lmbr1_prom_TelF
Typ35	AGCTGGGTAAGCTTCCATGA	Lmbr1_prom_TelR
Typ36	TGACTGAGAGGGGAACGTGA	Lmbr1_prom_CenF-2
Typ37	TTGGCAGGGGAGTGGTGTC	Lmbr1_prom_TelR-2
Typ49	GACCAAATTACTGGCTCACCA	i4_F
Typ50	TCGACAGAGGCTCTAAACTCG	i4_R
Typ51	TATGGCAGGCTGTTTGACAC	i5_F
Typ52	TCATAACTGGCAGCATTCTCA	i5_R
Typ53	TCACAGAAGAACAGCGCTAC	ZRS_CTCF F
Typ54	TCCTCTTGCCTGTGATTTCC	ZRS_Reg Tel F
Typ55	GCACTTCCTGAATCGCTCA	ZRS_Reg Tel F2
Typ56	TGTCCTGGTTTATGTCGCTT	ZRS_Reg Tel R
Typ57	ATTCTCCCAACACCTGCAAG	ZRS_CTCF cen R
Typ59	TGGTTCTAATCAAGGTATGGCA	i5 2_F
Typ60	AAGTCTCATAACTGGCAGCATT	i5 2_R
Typ61	TGCAAAAGGAGAAGGACGGA	i9_Cen_F
Typ62	AACAGTCCACACATTTGCCC	i9_Cen_2
Typ63	AGAAAACCACAAAGCAGCTCA	i9_Tel_F
Typ64	TGAAGAAATGCTGGAGTTTGGG	i9_Tel_R
Typ67	AGTATCCACAGCAAGCCACT	i4 2_F
Typ68	TGTGCCTCTACGTCACCATT	i4 2_R

Table 9. 3. Primers used for copy number analysis by qRT-PCR

NAME	SEQUENCE (5'--> 3')
Lmbr1_prom_del1F	CGGGGCTATACAGTCTCTCG
Lmbr1_prom_del1R	ATTCCCCGCCTCTTGTATCT
Lmbr1_prom_del2F	CCCTTCCATCCTCCTTTCAT
Lmbr1_prom_del2R	TCTAGGTGGCGTTTCTGTCC
qLmbr1ex6_FP	CAGGCCAACTGGAGTACACA
qLmbr1ex6_RP	TCTTGCGTTGCTGATTCTTG
qCTCF_Lmbr1_i4_F	CCACCTGCTGGTATCACAAA
qCTCF_Lmbr1_i4_R	TCGACAGAGGCTCTAAACTCG
qCTCF_Lmbr1_i5_F	CACCTAGCAGCCTCTGGAGT
qCTCF_Lmbr1_i5_R	GTCCCGCAAGCAGTTCTC
qLmbr1i5_FP	AAGGTATGGCAGGCTGTTTG
qLmbr1i5_RP	GAGGACTCCAGAGGCTGCTA
qZRS_CTCF_FP	CCACCCGTCACAGAAGAAC
qZRS_CTCF_RP1	ACTAGGGAGCACTGGTTTCT
qZRS_CTCF_RP2	AGAACTCTCTGCTCTCCCC
qZRS_Reg_F1	GCCAAGATCAAAACATGCCC
qZRS_Reg_R1	CGCAAACTCAGTCTGGTTCT
qZRS_Reg_F2	TCCTCTTGCCTGTGATTTCC
qZRS_Reg_R2	TGAGCGATTCAGGAAGTGC
qZRS_Reg_F3	TGAGCGATTCAGGAAGTGC
qZRS_Reg_R3	GAGCGTTCATTGGATTCTTTCA
qZRS_Reg_cen_F	AAACAGCATAGCCAGAGATGT
qZRS_Reg_cen_R	TAGTACCATGCGTGTGTGTG
qZRS_Reg_tel_F	GATGGCTGGATGGTTTGGAT
qZRS_Reg_tel_R	TCTTTGATTTGAAGTCCTGGCA
qCTCF_i9_F1	TGCCTCCCACACACTAGATT
qCTCF_i9_R1	GCACTGGAATACGCAGGAAC
qCTCF_i9_F2	ACCTTGGCTCTCAGACCTTT
qCTCF_i9_R2	TAAGACTGCAGGAGGATGGC
qCTCF_i9_F3	GAGTCACAGATCCTAAGTTTCCC
qCTCF_i9_R3	CAACAGGTGGCACTAGATCTC
qCTCF_i9_F4	AGCCGTCTCTCCAATCCTCT
qCTCF_i9_R4	TCGCATCACCTCTCATGTCC
qOut_i9_F1	TCTTAAGCTGGGCACGTTCC
qOut_i9_R1	TCCGTCCTTCTCCTTTTGCA
qOut_i9_F2	AGGGAAGCAGAAGAGGTAAGAA
qOut_i9_R2	GCTCACCAAGGCTCCAGAAG
qOut_i9_F3	CACCAGCAATTCCATCTCTAAGA
qOut_i9_R3	GGAGCCTTGGGTTTGAACAATC
qOut_i9_F4	TGATCCACAGTCTTCGGCTT
qOut_i9_R4	TGCAAGAGTGACTGAGGGAAG

Table 9. 4. Whole-Mount In Situ Hybridization buffers

BUFFER NAME	COMPOSITION
10x PBS -DEPC	1.37 M NaCl, 27 mM KCl, 100 mM Na2HPO4, 20mM KH2PO4, adjust pH to 7.4 with HCl, in DEPC- H2O, autoclave
4% PFA-PBS	Dissolve 40 mg/ml PFA in 1x PBS (DEPC), heat to 55°C until PFA is dissolved, adjust pH to 7.4 with HCl
Alkaline phosphatase buffer	0.02 M NaCl, 0.05 M MgCl2, 0.1% Tween 20, 0.1 M Tris pH 9.5, and 0.05% levamisole/tetramisole in H2O
Bleaching solution	6% H2O2/PBST
DEPC-H2O	0.1% DEPC in ddH2O
H1 hybridisation buffer	L1 with 0.1% tRNA and 0.05% heparin
H2 hybridisation buffer	H1 with 0.1% tRNA and 0.05% heparin and DIG probe
L1 buffer	50% Deionised formamide, 5x SSC, 1% SDS, 0.1% Tween-20 in DEPC; pH 4.5
L2 buffer	50% Deionised formamide, 2x SSC pH 4.5, 0.1% Tween-20 in DEPC; pH 4.5
L3 buffer	2x SSC pH 4.5, 0.1% Tween-20 in DEPC; pH 4.5
Proteinase K Buffer	20 mM Tris pH 7.0, 1 mM EDTA, in DEPC-H2O
TBST 1	140 mM NaCl, 2.7 mM KCl, 25 mM Tris-HCl, 1% Tween-20; pH 7.5
TBST 2	TBST with 0.1% Tween-20 and 0.05% levamisole/tetramisole
Blocking solution	TBST 1 with 2% calf-serum and 0.2% BSA
PBST	0.1 % Tween-20 in 1x PBS(DEPC)
RIPA buffer	Use DEPC treated reagents, 0.01% SDS, 0.15 M NaCl, 0.01% Nonidet-P40, 5 mg/ ml deoxycholate, 1 mM EDTA pH 8.0, 50 mM Tris pH 8.0, in DEPC-H2O
Rnase solution	0.1 M NaCl, 0.01 M Tris pH 7.5, 0.2% Tween-20, 100 µg/ml RNase A in H2O

9.3 Abbreviations

°C	Celcius degrees
µF	microF
µl	microliter
1x	1 Volume
3C	Chromosome Conformation Capture
3D	3-Dimensional
4C	Circular Chromosome Conformation Capture
Bid H_2O	Bidestilled water
bp	basepair
cDNA	complementary DNA
cen	centromeric
ChIP	Chromatin Immunoprecipitation
CREs	Cis-Regulatory Elements
CTCF	CCCCTC-factor
DEPC	Diethyl Pyrocarbonate
DSB	Double Strand Break
dsDNA	double-stranded DNA
E.coli	*Escerichia coli*
EtOH	Ethanol
gDNA	genomic DNA
GTF	General Trancriprion Factor
indel	insertions-deletions
kb	kilobase
KO	KnockOut
kV	kiloVolt
Mb	Megabase
mESC	mouse Embryonic Stem Cells
min	minute
mM	milliMolar
nt	nucleotide
ON	Overnight
PBS	Phosphate-Buffered Saline
PIC	PreInitiation Complex
Pol II	Polymerase II
rpm	rounds per minute
RT	Room Temperature
sec	second
Shh	Sonic Hedgehog
ssDNA	single-stranded DNA
TAD	Topologically Associating Domains
tel	telomeric
TF	Transcription Factor
TFBS	Transcription Factor Binding Site
TSS	Transcription Start Site

WISH	Whole Mount In-Situ Hybridization
ZPA	Zone of Polarizing Activity
ZRS	ZPA Regulatory Sequence
∞	infinite
sgRNA	single guide RNA
PCR	Polymerase Chair Reaction
ng	nanogram
mg	milligram
µg	microgram
ml	milliliter
LIF	Leukeumia Inhibiting Factor
mM	milliMolar
M	Molar
h	hour
PEG	PolyEthylene Glycol
mm	millimeter
pg	picogram
dNTPs	deoxyNucleotide TriPhosphate
w/v	weight per volume (g/L)

9.4 List of Figures

9.5 List of Tables

10 DECLARATION OF INDEPENDENT WORK

I hereby declare that I completed the doctoral thesis independently based on the stated resources and aids.
I have not applied for a doctoral degree elsewhere and do not have a corresponding doctoral degree.
I have not submitted the doctoral thesis, or parts of it, to another academic institution and the thesis has not been accepted or rejected.
I declare that I have acknowledged the Doctoral Degree Regulations which underlie the procedure of the Faculty of Mathematics and Natural Sciences of Humboldt-Universität zu Berlin, as amended on 27th June 2012.
Furthermore, I declare that no collaboration with commercial doctoral degree supervisors took place, and that the principles of Humboldt-Universität zu Berlin for ensuring good academic practice were abided by.

_____

Berlin, 20/06/2019

11 SCIENTIFIC PUBLICATION

Andrey G, Schöpflin R, Jerković I, Heinrich V, Ibrahim DM, **Paliou C**, Hochradel M, Timmermann B, Haas S, Vingron M, et al. 2017. Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding. *Genome Res* **27**: 223–233.

Despang A, Schöpflin R, Franke M, Ali S, Jerković I, **Paliou C**, Chan W-L, Timmermann B, Wittler L, Vingron M, et al. 2019. Functional dissection of TADs reveals non-essential and instructive roles in regulating gene expression. *bioRxiv*.

Kraft K, Geuer S, Will AJ, Chan W, **Paliou C**, Borschiwer M, Harabula I, Wittler L, Franke M, Ibrahim DM, et al. 2015. Deletions, inversions, duplications: Engineering of structural variants using CRISPR/Cas in mice. *Cell Rep* **10**: 833–839.

Kragesteen BK, Spielmann M, **Paliou C**, Heinrich V, Schöpflin R, Esposito A, Annunziatella C, Bianco S, Chiariello AM, Jerković I, et al. 2018. Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis. *Nat Genet* **50**: 1463–1473.

Paliou C, Andrey G. 2018. Large genomic insertion at the Shh locus results in hammer toes through enhancer adoption. *Proc Natl Acad Sci U S A* **115**: 10–12.

Paliou C, Guckelberger P, Schöpflin R, Heinrich V, Esposito A, Chiariello AM, Bianco S, Annunziatella C, Helmuth J, Haas S, et al. 2019. Preformed chromatin topology assists transcriptional robustness of Shh during limb development. *Proc Natl Acad Sci U S A*

12 ACKNOWLEDGEMENTS

I would like to thank Prof. Stefan Mundlos for the opportunity to conduct my PhD in his lab, a place with high research standards and an amazing scientific team. I consider myself lucky and I am grateful for everything I learnt from him and his people. I especially thank Prof. Ana Pombo who encouraged me to apply for PhD at this lab and was very helpful to me, whenever I needed it. Thanks to Prof. Leonie Ringrose for evaluating my dissertation and being part of my Doctoral degree committee.

Next, I want to thank everybody who joined forces with me, so that this PhD study becomes a beautiful publication. Especially, Guillaume, you have been an inspiration for me from the beginning to the very end of our common path in the Mundlos group. You were a supervisor, a mentor, a friend. Thank you for everything you taught me, including good cheese, fondue and whiskey sours, for having always my back and for the long chats about anything that was bothering us. I always did my best to make you proud of your first PhD student! My sweet Philine, you have been such a support in a difficult period for me. It would have been great to work together for longer period. Robert and Verena, thank you for the patience all the times you tried to "decode" what I wanted you to do with my data. I know it was not always easy ☺. Ivanaki, thanks for being such a respectful and ready-to-help colleague, but more importantly such a good friend. Your hug has been always a "shelter" for me. Sometimes, the Croatian nectar as well ;-) Fanisita, thank you for the positive vibes you always transmitted to me. I am lucky to have met you and shared so beautiful and unique experiences with you. You, your little cookie A. and all the Mediterranean girls, Chiara, Giulia, Lila, were my sunshine in the dark days. This office with a lot of laughter, sometimes crying, silence when needed, and amazing coziness felt like home! My lovely Chiara, our numerous in-depth conversations about work and life were so precious to me. We are the best self-reflecting duet ever! Many thanks to Magda and Rocío, who were only shortly my colleagues but that was enough for me to gain two more amazing friends in my life. If it weren't for all these great people who were and are the brain and the soul of the Mundlos group, these last five years would not have been so creative, productive and in countless instances so much fun. Thank you Sala, Alex, Bjørt, Darío, Mike, Alessa, Daniel and Katerina! Asita, Norbert and Ute, you keep the lab running so smoothly that makes our experiments feasible and successful. Thank you for your patience and dedication. My most special thanks to Martin for troubleshooting with me all scientific and non-scientific issues of the last four years. Thank you for being you, for being there, for your patience and support, for your soothing words, for reminding me who I am and making me a better scientist and person.

Last, but not least, I want to thank my family and friends back home who never left my side and always comforted me and supported me in following my dreams without fear but with passion. An

exclusive heartfelt thanks to my parents for the endless love, but also for the ethics and qualities they taught me. I would not be the person I am today without them! This study is dedicated to my grandfather who might have never understood Biology, but always supported my choice and taught me to "fly" as high as possible. Ένα ξεχωριστό από καρδιάς ευχαριστώ στους γονείς μου για την απέραντη αγάπη, αλλά και για τις αξίες και τα ιδανικά που με δίδαξαν. Δε θα ήμουν αυτό που είμαι σήμερα χωρίς αυτούς! Αυτή η διατριβή είναι αφιερωμένη στον παππού μου, που μπορεί να μην καταλαβαίνει Βιολογία, αλλά υποστήριξε την επιλογή μου στο έπακρο και με δίδαξε να «πετάω» όσο πιο ψηλά μπορώ! Σε ευχαριστώ!